ATTACHMENT A - SAMPLE SUBMITTAL MATERIALS

Including the following:

- AMP380 report
- Site ID forms
- Monitor ID form
- Sample MSA monitoring network design map
- Sample 2km radius map
- Sample 1/4 mile radius map
- Land use information

```
DATE  06/24/93                              EPA AEROMETRIC INFORMATION RETRIEVAL SYSTEM (AIRS)
AMP380                                                   AIR QUALITY SUBSYSTEM
                                                      SITE DESCRIPTION INVENTORY

EPA REGION: 02                                     STATE (72): PUERTO RICO

SITE ID: 72-001-9999    ADDRESS: 123 PAMS TEST LANE                          STATE/LOCAL ID:
CITY POPULATION :          1    CITY     (00000): NOT IN A CITY                   DIST CITY: 999 K     UTM ZONE : 19
AQCR POPULATION :  3,196,520    COUNTY     (001): ADJUNTAS CO                     DIFF GMT :           UTM NORTH:  2004041
                                DATE ESTABLISHED: 1992/01/01                      ELEV MSL :    999 M  UTM EAST :   728975
MET SITE:                       DATE TERMINATED :     /  /                        COMP SECT: NW        LATITUDE : +18:06:51
  SITE ID        :   -   -      DATE LAST UPDATE: 1993/06/24                                           LONGITUDE: -066:50:10
  DISTANCE SITE :        M      HQ EVAL DATE    :     /  /                        METHOD (MAP        ): MAP INTERPOLATION, VIA.DI
  DIRECTION SITE:               REGN EVAL DATE  :     /  /                        SCALE      (15840 ): 1:15,840
  TYPE SITE  (1): ON-SITE MET EQUIP                                               DATUM          (27): NAD 27
                                                                                  EST OF ACCURACY    : 5.00000SEC

AQCR                             (244): PUERTO RICO
CMSA LOCATED IN                 (0000): ** DESCRIPTION UNKNOWN **
LAND USE                           (1): RESIDENTIAL
LOCATION DESCRIPTION                  : TEST PAMS SITE FOR PUERTO RICO
LOCATION SETTING                   (1): URBAN AND CENTER CITY
MSA LOCATED IN                  (0000): NOT IN AN MSA
SITE COMMENT 1                        : SITE COMMENTS-1
SITE COMMENT 2                        : SITE COMMENTS-2
SUPPORTING AGENCY                (001): PUERTO RICO ENVRIONMENTAL QUALITY BOARD
URBAN AREA REPRESENTED          (5601): NEW YORK, NY-NORTHEASTERN NJ

TANGENT STREET:
STREET #  TRAFFIC FLOW  YR TRAFFIC  DIR  STREET NAME         TYPE ROAD
   1         999999        1990      NW  PAMS TEST MAIN ST   (1): ARTERIAL

SITE TYPE:
SITE LOCATION TYPE                                                           MSA OR CMSA REPRESENTED
(1): UPWIND BACKGROUND                                                        61          ********** DESCRIPTION UNKNOWN **********
(0): ********** DESCRIPTION UNKNOWN **********
(0): ********** DESCRIPTION UNKNOWN **********


PARAMETER          : 42602     DATE SAMPLING BEGAN: 1992/01/01    SITE CRITERIA MET      : Y     DATE SITE CRITERIA MET: 1992/01
POC                : 1         DATE SAMPLING ENDED:     /  /      REF MTHOD USED         : Y     REF METHOD USED DATE  : 1992/01
MONITOR TYPE       : 6         DATE TYPE EFFECTIVE: 1992/01/01    QA PLAN                :       QA EFFECTIVE DATE     :      /
REPORTING ORGANIZ  : 001       RO EFFECTIVE DATE  : 1992/01/01    ACTION TYPE            : APP   ACTION TYPE REASON    : 01
COLLECTING LAB     : 001       AUDIT DATE         :     /  /      MONITOR OPEN PATH NUM  :       PROJECT CLASS         :
ANALYZING LAB      : 001       PROBE LOCATION  (1): TOP OF BUILDING
UNRESTRIC AIR FLOW: Y          PROBE HEIGHT       :    9 M        HORIZONTAL DISTANCE    : 0 M   VERTICAL DISTANCE     : 1 M

DOMINANT SOURCE         (1): POINT
MEASUREMENT SCALE       (3): NEIGHBORHOOD - 500 M TO 4KM
MONITOR TYPE OBJECTIVE  (3): BACKGROUND
PAMS REQ SAMPLE FREQ    (S): SEASONAL
```

EPA AEROMETRIC INFORMATION RETRIEVAL SYSTEM (AIRS)
AIR QUALITY SUBSYSTEM
SITE DESCRIPTION INVENTORY

```
EPA REGION: 02                                    STATE (72): PUERTO RICO

SITE ID: 72-001-9999    PARAMETER: 42602    POC: 1
ROAD DESCRIPTION:
 STREET NUMBER          DISTANCE FROM ROAD
       1                      9999 M

MONITOR COMMENTS 1: COMMENTS-1 FOR NO2
MONITOR COMMENTS 2: COMMENTS-2 FOR NO2

PARAMETER          : 43104      DATE SAMPLING BEGAN: 1992/01/01    SITE CRITERIA MET      : Y      DATE SITE CRITERIA MET: 1992/01
POC                : 1          DATE SAMPLING ENDED:     /  /      REF MTHOD USED         : Y      REF METHOD USED DATE  : 1992/01
MONITOR TYPE       : 6          DATE TYPE EFFECTIVE: 1992/01/01    QA PLAN                :        QA EFFECTIVE DATE     :     /
REPORTING ORGANIZ  : 001        RO EFFECTIVE DATE  : 1992/01/01    ACTION TYPE            : APP    ACTION TYPE REASON    : 01
COLLECTING LAB     : 001        AUDIT DATE         :     /  /      MONITOR OPEN PATH NUM  :        PROJECT CLASS         :
ANALYZING LAB      : 001        PROBE LOCATION  (1): TOP OF BUILDING
UNRESTRIC AIR FLOW: N           PROBE HEIGHT       :    9 M        HORIZONTAL DISTANCE    : 0 M    VERTICAL DISTANCE     : 1 M

DOMINANT SOURCE         (1): POINT
MEASUREMENT SCALE       (3): NEIGHBORHOOD - 500 M TO 4KM
MONITOR TYPE OBJECTIVE  (3): BACKGROUND
PAMS REQ SAMPLE FREQ    (S): SEASONAL

PROBE OBSTRUCTION:
 TYPE OBSTRUCTION           DIRECTION FROM MONITOR   DISTANCE FROM INLET   HEIGHT OF OBS
 (1): BUILDINGS                       NW                     99 M               99 M

ROAD DESCRIPTION:
 STREET NUMBER          DISTANCE FROM ROAD
       1                        99 M

MONITOR COMMENTS 1: MONITOR COMMENTS-1 FOR VOC
MONITOR COMMENTS 2: MONITOR COMMENTS-2 FOR VOC

PARAMETER          : 43502      DATE SAMPLING BEGAN: 1992/01/01    SITE CRITERIA MET      : Y      DATE SITE CRITERIA MET: 1992/01
POC                : 1          DATE SAMPLING ENDED:     /  /      REF MTHOD USED         : Y      REF METHOD USED DATE  : 1992/01
MONITOR TYPE       : 6          DATE TYPE EFFECTIVE: 1992/01/01    QA PLAN                :        QA EFFECTIVE DATE     :     /
REPORTING ORGANIZ  : 001        RO EFFECTIVE DATE  : 1992/01/01    ACTION TYPE            : AAP    ACTION TYPE REASON    : 01
COLLECTING LAB     : 001        AUDIT DATE         :     /  /      MONITOR OPEN PATH NUM  :        PROJECT CLASS         :
ANALYZING LAB      : 001        PROBE LOCATION  (1): TOP OF BUILDING
UNRESTRIC AIR FLOW: Y           PROBE HEIGHT       :    9 M        HORIZONTAL DISTANCE    : 1 M    VERTICAL DISTANCE     : 0 M

DOMINANT SOURCE         (1): POINT
MEASUREMENT SCALE       (3): NEIGHBORHOOD - 500 M TO 4KM
MONITOR TYPE OBJECTIVE  (3): BACKGROUND
PAMS REQ SAMPLE FREQ    (S): SEASONAL

ROAD DESCRIPTION:
 STREET NUMBER          DISTANCE FROM ROAD
       1                      9999 M

MONITOR COMMENTS 1: MONITOR COMMENTS-1 FOR FORMALDEHYDE
MONITOR COMMENTS 2: MONITOR COMMENTS-2 FOR FORMALDEHYDE
```

```
DATE  06/24/93                               EPA AEROMETRIC INFORMATION RETRIEVAL SYSTEM (AIRS)
AMP380                                                   AIR QUALITY SUBSYSTEM
                                                      SITE DESCRIPTION INVENTORY

EPA REGION: 02                                      STATE (72): PUERTO RICO

SITE ID: 72-001-9999
PARAMETER           : 43503     DATE SAMPLING BEGAN: 1992/01/01    SITE CRITERIA MET       : Y     DATE SITE CRITERIA MET: 1992/01
POC                 : 1         DATE SAMPLING ENDED:     /  /      REF MTHOD USED          : Y     REF METHOD USED DATE  : 1992/01
MONITOR TYPE        : 6         DATE TYPE EFFECTIVE: 1992/01/01    QA PLAN                 :       QA EFFECTIVE DATE     :     /
REPORTING ORGANIZ   : 001       RO EFFECTIVE DATE  : 1992/01/01    ACTION TYPE             : APP   ACTION TYPE REASON    : 01
COLLECTING LAB      : 001       AUDIT DATE         :     /  /      MONITOR OPEN PATH NUM   :       PROJECT CLASS         :
ANALYZING LAB       : 001       PROBE LOCATION  (1): TOP OF BUILDING
UNRESTRIC AIR FLOW: Y           PROBE HEIGHT       :    9 M        HORIZONTAL DISTANCE     : 1 M   VERTICAL DISTANCE     : 0 M

DOMINANT SOURCE         (1): POINT
MEASUREMENT SCALE       (3): NEIGHBORHOOD - 500 M TO 4KM
MONITOR TYPE OBJECTIVE  (3): BACKGROUND
PAMS REQ SAMPLE FREQ    (S): SEASONAL

ROAD DESCRIPTION:
 STREET NUMBER          DISTANCE FROM ROAD
       1                      9999 M

MONITOR COMMENTS 1: COMMENTS-1 FOR PARAMETER 43503
MONITOR COMMENTS 2: COMMENTS-2 FOR PARAMETER 43503

PARAMETER           : 43551     DATE SAMPLING BEGAN: 1992/01/01    SITE CRITERIA MET       : Y     DATE SITE CRITERIA MET: 1992/01
POC                 : 1         DATE SAMPLING ENDED:     /  /      REF MTHOD USED          : Y     REF METHOD USED DATE  : 1992/01
MONITOR TYPE        : 6         DATE TYPE EFFECTIVE: 1992/01/01    QA PLAN                 :       QA EFFECTIVE DATE     :     /
REPORTING ORGANIZ   : 001       RO EFFECTIVE DATE  : 1992/01/01    ACTION TYPE             : APP   ACTION TYPE REASON    : 01
COLLECTING LAB      : 001       AUDIT DATE         :     /  /      MONITOR OPEN PATH NUM   :       PROJECT CLASS         :
ANALYZING LAB       : 001       PROBE LOCATION  (1): TOP OF BUILDING
UNRESTRIC AIR FLOW: Y           PROBE HEIGHT       :    9 M        HORIZONTAL DISTANCE     : 0 M   VERTICAL DISTANCE     : 1 M

DOMINANT SOURCE         (1): POINT
MEASUREMENT SCALE       (3): NEIGHBORHOOD - 500 M TO 4KM
MONITOR TYPE OBJECTIVE  (3): BACKGROUND
PAMS REQ SAMPLE FREQ    (S): SEASONAL

ROAD DESCRIPTION:
 STREET NUMBER          DISTANCE FROM ROAD
       1                      9999 M

MONITOR COMMENTS 1: MONITOR COMMENTS-1 FOR ACETONE
MONITOR COMMENTS 2: MONITOR COMMENTS-2 FOR ACETONE

PARAMETER           : 44201     DATE SAMPLING BEGAN: 1992/01/01    SITE CRITERIA MET       : Y     DATE SITE CRITERIA MET: 1992/03
POC                 : 1         DATE SAMPLING ENDED:     /  /      REF MTHOD USED          : Y     REF METHOD USED DATE  : 1992/01
MONITOR TYPE        : 6         DATE TYPE EFFECTIVE: 1992/01/01    QA PLAN                 :       QA EFFECTIVE DATE     :     /
REPORTING ORGANIZ   : 001       RO EFFECTIVE DATE  : 1992/01/01    ACTION TYPE             : APP   ACTION TYPE REASON    : 01
COLLECTING LAB      : 001       AUDIT DATE         :     /  /      MONITOR OPEN PATH NUM   :       PROJECT CLASS         :
ANALYZING LAB       : 001       PROBE LOCATION  (1): TOP OF BUILDING
UNRESTRIC AIR FLOW: Y           PROBE HEIGHT       :    9 M        HORIZONTAL DISTANCE     : 0 M   VERTICAL DISTANCE     : 1 M
```

EPA AERONETRIC INFORMATION RETRIEVAL SYSTEM (AIRS)
AIR QUALITY SUBSYSTEM
SITE DESCRIPTION INVENTORY

```
EPA REGION: 02                                STATE (72): PUERTO RICO

SITE ID: 72-001-9999     PARAMETER: 44201    POC: 1
DOMINANT SOURCE          (1): POINT
MEASUREMENT SCALE        (3): NEIGHBORHOOD - 500 M TO 4KM
MONITOR TYPE OBJECTIVE   (3): BACKGROUND
PAMS REQ SAMPLE FREQ     (S): SEASONAL

ROAD DESCRIPTION:
 STREET NUMBER           DISTANCE FROM ROAD
      1                        999 M

MONITOR COMMENTS 1: COMMENTS-1 FOR OZONE
MONITOR COMMENTS 2: COMMENTS-2 FOR OZONE

PARAMETER           : 61102    DATE SAMPLING BEGAN: 1992/01/01   SITE CRITERIA MET       : Y     DATE SITE CRITERIA MET: 1992/01
POC                 : 1        DATE SAMPLING ENDED:     /  /     REF MTHOD USED          : Y     REF METHOD USED DATE  : 1992/01
MONITOR TYPE        : 6        DATE TYPE EFFECTIVE: 1992/01/01   QA PLAN                 :       QA EFFECTIVE DATE     :     /
REPORTING ORGANIZ   : 001      RO EFFECTIVE DATE  : 1992/01/01   ACTION TYPE             : APP   ACTION TYPE REASON    : 01
COLLECTING LAB      : 001      AUDIT DATE         :     /  /     MONITOR OPEN PATH NUM   :       PROJECT CLASS         :
ANALYZING LAB       : 001      PROBE LOCATION  (1): TOP OF BUILDING
UNRESTRIC AIR FLOW: Y          PROBE HEIGHT       :    9 M       HORIZONTAL DISTANCE     : 1 M   VERTICAL DISTANCE     : 0 M

DOMINANT SOURCE          (1): POINT
PAMS REQ SAMPLE FREQ     (S): SEASONAL

ROAD DESCRIPTION:
 STREET NUMBER           DISTANCE FROM ROAD
      1                         9 M

MONITOR COMMENTS 1: COMMENTS-1 FOR WIND DIRECTION
MONITOR COMMENTS 2: COMMENTS-2 FOR WIND DIRECTION

PARAMETER           : 62101    DATE SAMPLING BEGAN: 1992/01/01   SITE CRITERIA MET       : Y     DATE SITE CRITERIA MET: 1992/01
POC                 : 1        DATE SAMPLING ENDED:     /  /     REF MTHOD USED          : Y     REF METHOD USED DATE  : 1992/01
MONITOR TYPE        : 6        DATE TYPE EFFECTIVE: 1992/01/01   QA PLAN                 :       QA EFFECTIVE DATE     :     /
REPORTING ORGANIZ   : 001      RO EFFECTIVE DATE  : 1992/01/01   ACTION TYPE             : APP   ACTION TYPE REASON    : 01
COLLECTING LAB      : 001      AUDIT DATE         :     /  /     MONITOR OPEN PATH NUM   :       PROJECT CLASS         :
ANALYZING LAB       : 001      PROBE LOCATION  (1): TOP OF BUILDING
UNRESTRIC AIR FLOW: Y          PROBE HEIGHT       :    9 M       HORIZONTAL DISTANCE     : 1 M   VERTICAL DISTANCE     : 0 M

DOMINANT SOURCE          (1): POINT
MEASUREMENT SCALE        (3): NEIGHBORHOOD - 500 M TO 4KM
MONITOR TYPE OBJECTIVE   (3): BACKGROUND
PAMS REQ SAMPLE FREQ     (S): SEASONAL

ROAD DESCRIPTION:
 STREET NUMBER           DISTANCE FROM ROAD
      1                       9999 M

MONITOR COMMENTS 1: COMMENTS-1 FOR TEMPERATURE
MONITOR COMMENTS 2: COMMENTS-2 FOR TEMPERATURE
```

EPA AEROMETRIC INFORMATION RETRIEVAL SYSTEM (AIRS)
AIR QUALITY SUBSYSTEM
SITE DESCRIPTION INVENTORY

```
EPA REGION: 02                              STATE (72): PUERTO RICO

SITE ID: 72-001-9999
PARAMETER           : 62201      DATE SAMPLING BEGAN: 1992/01/01     SITE CRITERIA MET       : Y      DATE SITE CRITERIA MET: 1992/01
POC                 : 1          DATE SAMPLING ENDED:    /  /        REF MTHOD USED          : Y      REF METHOD USED DATE  : 1992/01
MONITOR TYPE        : 6          DATE TYPE EFFECTIVE: 1992/01/01     QA PLAN                 :        QA EFFECTIVE DATE     :     /
REPORTING ORGANIZ   : 001        RO EFFECTIVE DATE  : 1992/01/01     ACTION TYPE             : APP    ACTION TYPE REASON    : 01
COLLECTING LAB      : 001        AUDIT DATE         :    /  /        MONITOR OPEN PATH NUM   :        PROJECT CLASS         :
ANALYZING LAB       : 001        PROBE LOCATION  (1): TOP OF BUILDING
UNRESTRIC AIR FLOW: Y            PROBE HEIGHT       :    9 M         HORIZONTAL DISTANCE     : 0 M    VERTICAL DISTANCE     : 1 M

DOMINANT SOURCE          (1): POINT
MEASUREMENT SCALE        (3): NEIGHBORHOOD - 500 M TO 4KM
MONITOR TYPE OBJECTIVE   (3): BACKGROUND
PAMS REQ SAMPLE FREQ     (S): SEASONAL

ROAD DESCRIPTION:
 STREET NUMBER            DISTANCE FROM ROAD
      1                         9999 M

MONITOR COMMENTS 1: COMMENTS-1 FOR HUMIDITY
MONITOR COMMENTS 2: COMMENTS-2 FOR HUMIDITY

PARAMETER           : 63301      DATE SAMPLING BEGAN: 1992/01/01     SITE CRITERIA MET       : Y      DATE SITE CRITERIA MET: 1992/01
POC                 : 1          DATE SAMPLING ENDED:    /  /        REF MTHOD USED          : Y      REF METHOD USED DATE  : 1992/01
MONITOR TYPE        : 6          DATE TYPE EFFECTIVE: 1992/01/01     QA PLAN                 :        QA EFFECTIVE DATE     :     /
REPORTING ORGANIZ   : 001        RO EFFECTIVE DATE  : 1992/01/01     ACTION TYPE             : APP    ACTION TYPE REASON    : 01
COLLECTING LAB      : 001        AUDIT DATE         :    /  /        MONITOR OPEN PATH NUM   :        PROJECT CLASS         :
ANALYZING LAB       : 001        PROBE LOCATION  (1): TOP OF BUILDING
UNRESTRIC AIR FLOW: N            PROBE HEIGHT       :    9 M         HORIZONTAL DISTANCE     : 0 M    VERTICAL DISTANCE     : 1 M

DOMINANT SOURCE          (1): POINT
MEASUREMENT SCALE        (3): NEIGHBORHOOD - 500 M TO 4KM
MONITOR TYPE OBJECTIVE   (3): BACKGROUND
PAMS REQ SAMPLE FREQ     (S): SEASONAL

ROAD DESCRIPTION:
 STREET NUMBER            DISTANCE FROM ROAD
      1                         9999 M

MONITOR COMMENTS 1: COMMENTS-1 FOR SOLAR RADIATION
MONITOR COMMENTS 2: COMMENTS-2 FOR SOLAR RADIATION

PARAMETER           : 64101      DATE SAMPLING BEGAN: 1992/01/01     SITE CRITERIA MET       : Y      DATE SITE CRITERIA MET: 1992/01
POC                 : 1          DATE SAMPLING ENDED:    /  /        REF MTHOD USED          : Y      REF METHOD USED DATE  : 1992/01
MONITOR TYPE        : 6          DATE TYPE EFFECTIVE: 1992/01/01     QA PLAN                 :        QA EFFECTIVE DATE     :     /
REPORTING ORGANIZ   : 001        RO EFFECTIVE DATE  : 1992/01/01     ACTION TYPE             : APP    ACTION TYPE REASON    : 01
COLLECTING LAB      : 001        AUDIT DATE         :    /  /        MONITOR OPEN PATH NUM   :        PROJECT CLASS         :
ANALYZING LAB       : 001        PROBE LOCATION  (1): TOP OF BUILDING
UNRESTRIC AIR FLOW: N            PROBE HEIGHT       :    9 M         HORIZONTAL DISTANCE     : 0 M    VERTICAL DISTANCE     : 1 M
```

DATE 06/24/93
AMP380

EPA AEROMETRIC INFORMATION RETRIEVAL SYSTEM (AIRS)
AIR QUALITY SUBSYSTEM
SITE DESCRIPTION INVENTORY

EPA REGION: 02   STATE (72): PUERTO RICO

SITE ID: 72-001-9999   PARAMETER: 64101   POC: 1
DOMINANT SOURCE   (1): POINT
MEASUREMENT SCALE   (3): NEIGHBORHOOD - 500 M TO 4KM
MONITOR TYPE OBJECTIVE   (3): BACKGROUND
PAMS REQ SAMPLE FREQ   (S): SEASONAL

PROBE OBSTRUCTION:

TYPE OBSTRUCTION	DIRECTION FROM MONITOR	DISTANCE FROM INLET	HEIGHT OF OBS
(2): TREES/BRUSH	NW	9 M	9 M
(3): RIDGES	NE	9 M	9 M

ROAD DESCRIPTION:

STREET NUMBER	DISTANCE FROM ROAD
1	9999 M

MONITOR COMMENTS 1: COMMENTS-1 FOR BAROMETRIC PRESSURE
MONITOR COMMENTS 2: COMMENTS-2 FOR BAROMETRIC PRESSURE

OMB No. 2060-0084
Approval Expires 01/31/96

EPA
AEROMETRIC INFORMATION RETRIEVAL SYSTEM
RESEARCH TRIANGLE PARK, NC 27711

SITE IDENTIFICATION FORM

FORM COMPLETED BY ____________________

DATE ____________________

STATE ____________________ CITY ____________________ COUNTY ____________________

TRAN TYPE	STATE	COUNTY	SITE	SUPPORT AGENCY	STREET ADDRESS	SITE ESTABLISHED	SITE TERMINATED	HQ EVAL DATE	REG EVAL DATE	TRANS ID	ACT
1	2–3	4–6	7–10	11–13	14–53	54–59	60–65	66–71	72–77	78–79	80
A										1	

TRAN TYPE	STATE	COUNTY	SITE	CITY	AQCR	MSA	AIR BASIN	ELEVATION MSL	COMPASS SECTOR	DISTANCE CITY	LOC SET	LAND USE	DIF GMT	TMS	MET SITE ID	DISTANCE SITE	DIRECT SITE	STATE OR LOCAL ID	TRANS ID	ACT
1	2–3	4–6	7–10	11–15	16–18	19–22	23–24	25–29	30–31	32–33	34	35	36–37	38	39–48	49–53	54–55	56–65	78–79	80
A																			2	

TRAN TYPE	STATE	COUNTY	SITE	SEQ NO	ST NO	STREET NAME	TYPE RD	TRAFFIC FLOW	TRAF YR	DIR ST	SLT	MSA REP	CMSA REP	PATH #	DIR TRANS	BEAM LENGTH	TRAN HT	RECV HT	MIN HT	MAX HT	LUP	TRANS ID	ACT
A																						3	

TRAN TYPE	STATE	COUNTY	SITE	SEQ NO	ST NO	STREET NAME	TYPE RD	TRAFFIC FLOW	TRAF YR	DIR ST	SLT	MSA REP	CMSA REP	PATH #	DIR TRANS	BEAM LENGTH	TRAN HT	RECV HT	MIN HT	MAX HT	LUP	TRANS ID	ACT
A																						3	

TRAN TYPE	STATE	COUNTY	SITE	SEQ NO	ST NO	STREET NAME	TYPE RD	TRAFFIC FLOW	TRAF YR	DIR ST	SLT	MSA REP	CMSA REP	PATH #	DIR TRANS	BEAM LENGTH	TRAN HT	RECV HT	MIN HT	MAX HT	LUP	TRANS ID	ACT
A																						3	

NOTES:

Card	Field	Meaning
Card A1–A3	TRAN TYPE:	Transaction Type
Card A1–A3	TRANS ID:	Transaction-ID
Card A1	DATE:	YYMMDD Format
Card A1	HQ EVAL:	Headquarters Evaluation
Card A1	ACT:	Action Code (Delete = D)
Card A1–A3	ACT:	Action Code (Insert = I, Modify = M)
Card A2	DIF GMT:	Difference Greenwich Mean Time
Card A2	LOC SET:	Location Setting
Card A2	TMS:	Type Met Site
Card A2	DIRECT SITE:	Direction to Site
Card A3	SEQ NO:	Sequence Number
Card A3	ST NO:	Street Number
Card A3	TYPE RD:	Type Road
Card A3	TRAF YR:	Year of Traffic Flow
Card A3	DIR ST:	Direction to Street
Card A3	SLT:	Site Location Type
Card A3	MSA REP:	MSA Represented
Card A3	CMSA REP:	CMSA Represented
Card A3	PATH #:	Open Path Number
Card A3	DIR TRANS:	Direction to Transmitter
Card A3	TRANS HT:	Transmitter Height
Card A3	RECV HT:	Receiver Height
Card A3	MIN HT:	Min Beam Height
Card A3	MAX HT:	Max Beam Height
Card A3	LUP:	Land Use Under Path

Session 3

OMB No. 2060-0084
Approval Expires 01/31/98

EPA
AEROMETRIC INFORMATION RETRIEVAL SYSTEM
RESEARCH TRIANGLE PARK, NC 27711

SITE IDENTIFICATION FORM

FORM COMPLETED BY ____________________

DATE ____________________

STATE ____________________ CITY ____________________ COUNTY ____________________

TRAN TYPE	STATE	COUNTY	SITE	UNIT ZONE	UTM - EASTING	UTM - NORTHING	SIGN	LATITUDE DEG	LATITUDE MIN	LATITUDE SEC	S S S S	SIGN	LONGITUDE DEG	LONGITUDE MIN	LONGITUDE SEC	S S S S	METHOD OF DETERMINATION	DATUM	SCALE	ESTIMATE OF ACCURACY	TRANS ID	ACT
1	2–3	4–6	7–10	11–12	13–20	21–27	28	29–30	31–32	33–34	35–38	39	40–42	43–44	45–46	47–50	51–60	61–62	63–68	69–78	79	80
A																					4	

TRAN TYPE	STATE	COUNTY	SITE	LOCATION DESCRIPTION	TRANS ID	ACT
1	2–3	4–6	7–10	11–78	79	80
A					5	

TRAN TYPE	STATE	COUNTY	SITE	SITE COMMENTS (LINE 1)	TRANS ID	ACT
1	2–3	4–6	7–10	11–78	79	80
A					6	

TRAN TYPE	STATE	COUNTY	SITE	SITE COMMENTS (LINE 2)	TRANS ID	ACT
1	2–3	4–6	7–10	11–78	79	80
A					7	

NOTES:

Card A4	SIGN:	"+" or "-"
Card A4	SSSS:	Decimal Portion of Field
Card A4–A7	TRAN TYPE:	Transaction Type
Card A4–A7	TRANS ID:	Transaction-ID
Card A4–A7	ACT:	Action Code (Insert = I, Modify = M)

OMB No. 2060-0064
Approval Expires 01/31/96

EPA
AEROMETRIC INFORMATION RETRIEVAL SYSTEM
RESEARCH TRIANGLE PARK, NC 27711

MONITOR IDENTIFICATION FORM

FORM COMPLETED BY ____________________

DATE ____________________

STATE ____________________ CITY ____________________ COUNTY ____________________

Card 1 (TRAN TYPE "F" in column 1, "1" in column 79)

TRAN TYPE (1) | MONITOR-ID: STATE (2–3), COUNTY (4–6), SITE (7–10), PARAMETER (11–15) | POC (16) | MON TYPE (17) | DATE TYPE EFF (18–23) | ANALYZE LAB (24–26) | COLLECT LAB (27–29) | REPORT ORG (30–32) | REPORT ORG EFFECTIVE DATE (33–38) | DATE: SAMPLING BEGAN (39–44), SAMPLING ENDED (45–50), LAST EPA AUDIT (51–56) | PROJECT CLASS (57–58) | ACTION TYPE (59–61) | ACT REASON (62–63) | DOM SRC (64) | M SCALE (65) | MTO (66) | NONATTAIN AREA (67–71) | URBAN AREA REP (72–75) | PRSF (76) | MOPN (77) | TRANS ID (78–79) | ACT (80)

Card 2 (TRAN TYPE "F" in column 1, "2" in column 79)

TRAN TYPE (1) | MONITOR-ID: STATE (2–3), COUNTY (4–6), SITE (7–10), PARAMETER (11–15) | POC (16) | MSN (17) | STREET INFORMATION: DISTANCE (18–21), MSN (22), DISTANCE (23–26), MSN (27), DISTANCE (28–31) | RM USED (32) | RM USED DATE Y Y M M (33–36) | STCR (37) | DATE SITE CRITERIA MET Y Y M M (38–41) | QA PLAN (42) | QA EFFECTIVE DATE Y Y M M (43–46) | PLC (47) | PROBE HEIGHT (48–50) | H DIST (51) | V DIST (52) | URAF (53) | OBSTRUCTION DESCRIPTION: TYPE OB (54), DIR MON (55–56), DIST INL (57–58), HT OBS (59–60), TYPE OB (61), DIR MON (62–63), DIST INL (64–65), HT OBS (66–67), TYPE OB (68), DIR MON (69–70), DIST INL (71–72), HT OBS (73–74) | TRANS ID (78–79) | ACT (80)

Card 3 (TRAN TYPE "F" in column 1, "3" in column 79)

TRAN TYPE (1) | MONITOR-ID: STATE (2–3), COUNTY (4–6), SITE (7–10), PARAMETER (11–15) | POC (16) | MON AREA (17–20) | WS (21) | SFLAG (22) | NUMBER OF SCHEDULED SAMPLES PER MONTH: JAN (23–24), FEB (25–26), MAR (27–28), APR (29–30), MAY (31–32), JUN (33–34), JUL (35–36), AUG (37–38), SEP (39–40), OCT (41–42), NOV (43–44), DEC (45–46) | RSF (47) | RSF EFF DATE (48–53) | STSS (54) | STCD (55–60) | FEC (61) | TRANS ID (78–79) | ACT (80)

Card 4 (TRAN TYPE "F" in column 1, "4" in column 79)

TRAN TYPE (1) | MONITOR-ID: STATE (2–3), COUNTY (4–6), SITE (7–10), PARAMETER (11–15) | POC (16) | MONITOR COMMENTS (LINE 1) (17–77) | TRANS ID (78–79) | ACT (80)

Card 5 (TRAN TYPE "F" in column 1, "5" in column 79)

TRAN TYPE (1) | MONITOR-ID: STATE (2–3), COUNTY (4–6), SITE (7–10), PARAMETER (11–15) | POC (16) | MONITOR COMMENTS (LINE 2) (17–77) | TRANS ID (78–79) | ACT (80)

NOTES:

Card	Field	Meaning
Card F1–F5	TRAN TYPE:	Transaction Type
Card F1–F5	TRANS ID:	Transaction-ID
Card F1–F5	POC:	Parameter Occurrence Code
Card F1	MON TYPE:	Monitor Type
Card F1	DATE TYPE EFF:	Date Type Effective
Card F1	REPORT ORG:	Reporting Organization
Card F1, F3	DATE:	YYMMDD Format
Card F1	AT REASON:	Action Type Reason
Card F1	M SCALE:	Measurement Scale
Card F1	MTO:	Monitor Type Objective
Card F1	DOM SRC:	Dominant Source
Card F1	PRSF:	Pams Req Sampling Freq
Card F1	MOPN:	Monitor Open Path Number
Card F1	ACT:	Action Code (Delete = D)
Card F1–F5	ACT:	Action Code (Insert = I, Modify = M)
Card F2	MSN:	Monitor Street Number
Card F2	DISTANCE:	Distance from Road
Card F2	RM:	Reference Method
Card F2	STCR:	Siting Criteria Met
Card F2	QA PLAN:	Quality Assurance Plan
Card F2	QA EFFECTIVE DATE:	QA Plan Effective Date
Card F2	PLC:	Probe Location Code
Card F2	H DIST:	Horizontal Distance
Card F2	V DIST:	Vertical Distance
Card F2	URAF:	Unrestricted Air Flow
Card F2	TYPE OB:	Type Obstruction
Card F2	DIR MON:	Direction from Monitor
Card F2	DIST INL:	Distance from Inlet
Card F3	HT OBS:	Height of Obstruction
Card F3	MON AREA:	Monitoring Area
Card F3	WS:	Worst Site
Card F3	SFLAG:	Surrogate Flag
Card F3	RSF:	Required Sampling Frequency
Card F3	RSF EFF DATE:	Required Sampling Frequency Effective Date
Card F3	STSS:	Short Term Scheduled Satisfied
Card F3	SCTD:	Short Term Completion Date
Card F3	FEC:	First Exceedance Correction

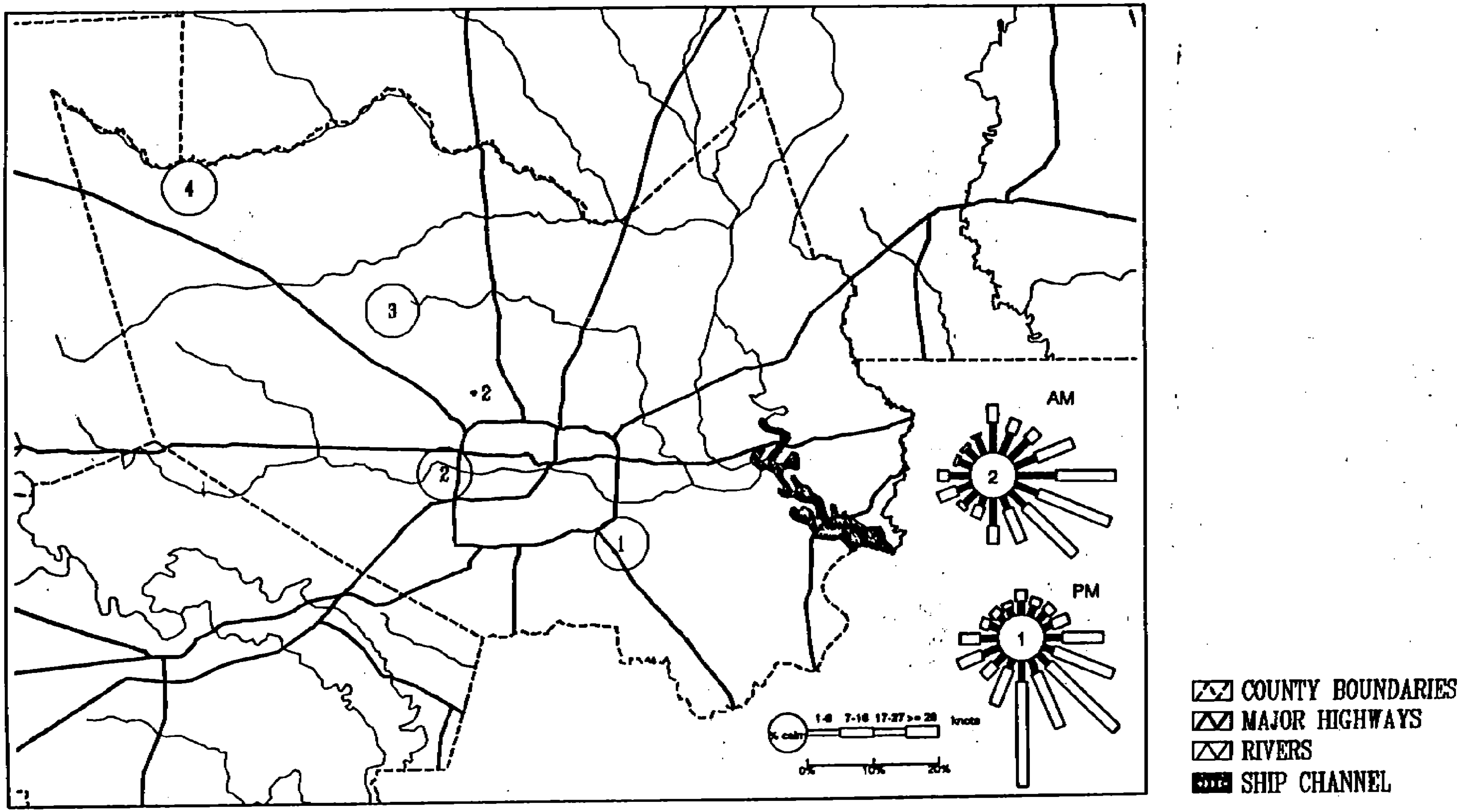


Figure A-1. Sample MSA Network Design Map

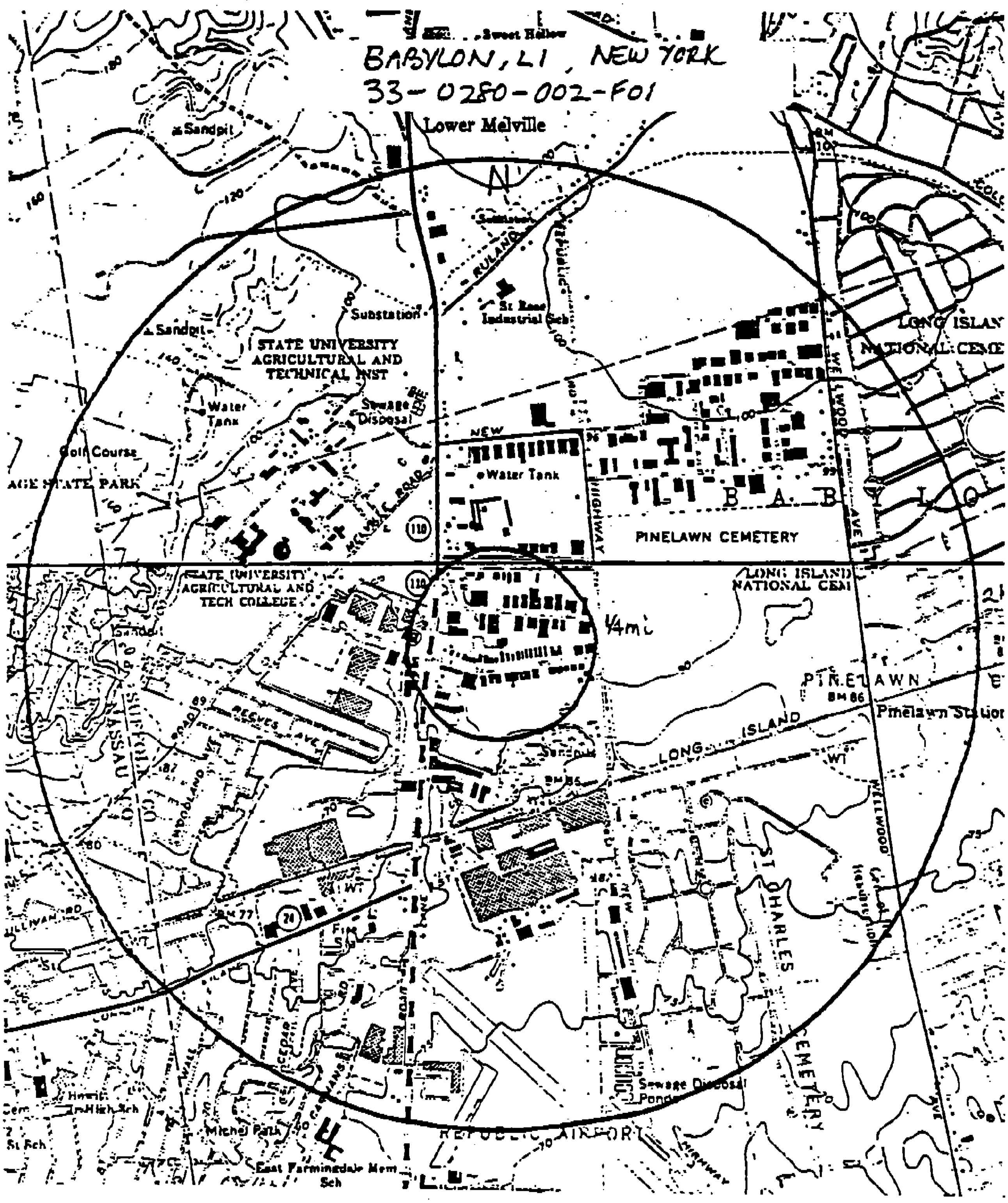


Figure A-2. Sample 2 Km Radius Map

Sketch a map to document the environment within a 1/4 mile radius of the site.
Include the following information on the drawing where applicable.

PAMS Location on Drawing
Roadways with Names (paved and unpaved)
Parking Areas (paved and unpaved)
Stationary Sources
Buildings (number of stories)
Undeveloped Land (ground cover)
Tree Lines or Clusters
Small Area Sources (dry cleaners, gas stations, etc.)
Residences
Trailer Parks
Recreation Parks
Recreation Fields
Railroad Yards
Bodies of Water
North Direction

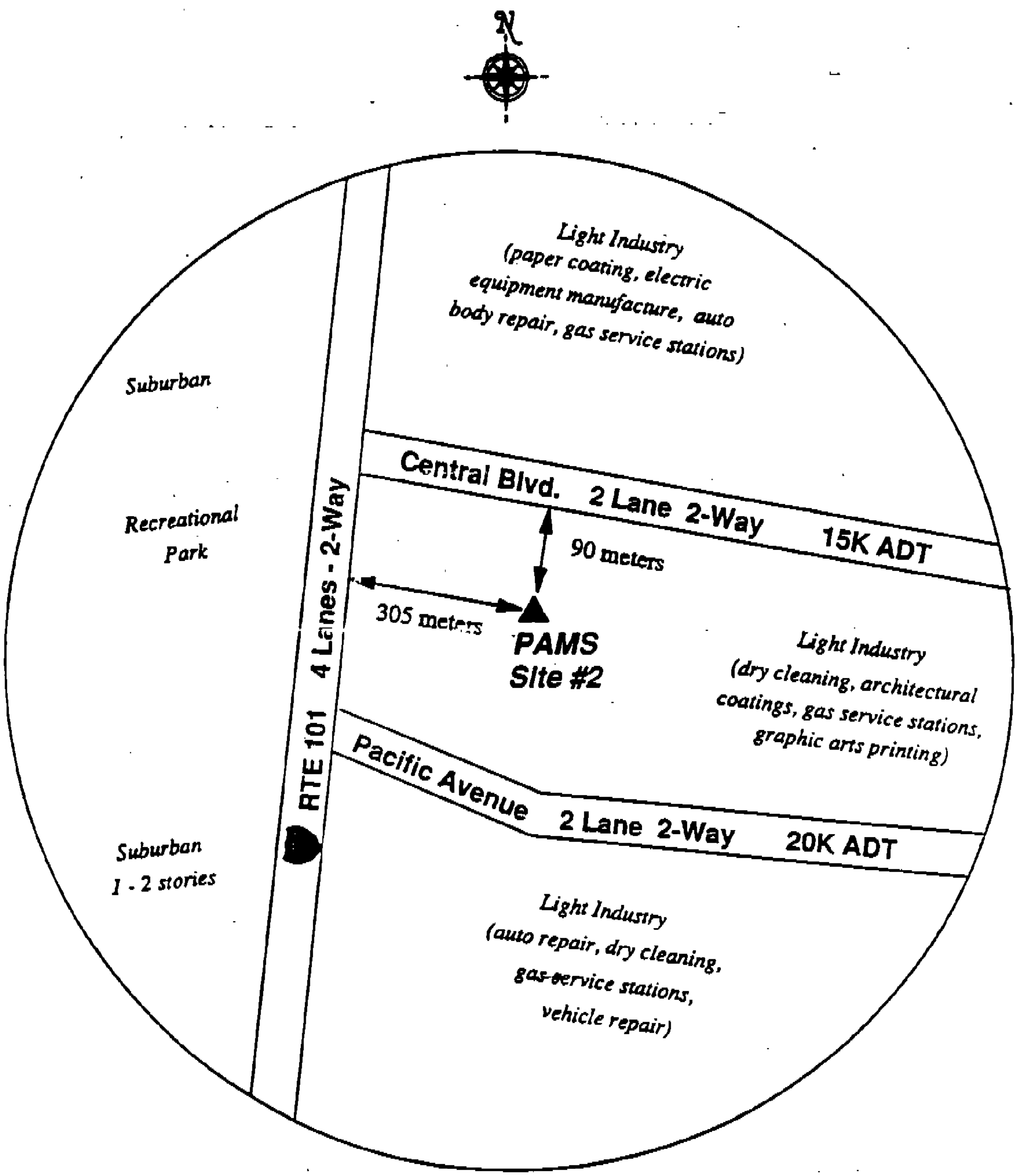


Figure A 3. Sample 1/4 Mile Radius Map

LAND USE INFORMATION

As noted in Section 2 of this document, land use information is quite useful in determining the appropriate locations for ambient monitoring stations.

Land use information availability varies across the country. Differing State and local land use regulations generate similar differences in land use data. Federal data sources also vary in the availability of detailed data on electronic media.

Land use is defined differently by the many public and private data vendors, as well as data customers. Land use can simply describe the degree of urbanization of an area and the nature of this man-made development, or it can describe the purpose served by the development (*e.g.*, residential, commercial, institutional, open space) as well as the transportation systems serving these areas.

The primary land use data sources in the federal government are the United States Geological Survey (USGS) and the Bureau of the Census. The USGS produces topographic maps on paper, electronic tape/disk, and optical disk. These maps are relatively inexpensive and dense with information, but land use information must be distilled from the maps' structure types, development densities, and transportation networks rather than simply being read as pre-marked zones. Topographic maps are available in a variety of scales and media for many areas of the country. Aerial photographs are available at 1:24,000 scale. These photographs offer literal images of the surface features photographed from aircraft. Trained readers can classify types of development and vegetation from these monochrome prints. Pilot projects have yielded orthophotoquad maps on optical disk for a few areas. The Bureau of the Census provides demographic and socioeconomic data capable of supporting land use analysis.

State, regional, and local governments maintain varying amounts of land use data. Data maintained often vary with the planning requirements mandated by State or local legislation. An advantage of sub-state level data is the presence of projected land use maps. Comprehensive development plans often include current land use and projected land use, the latter typically being in the form of zoning maps which reflect policies steering development.

With the development of satellite technology and image processing computer systems, remote sensing has become a viable source of land cover, topographic, and interpretive data such as land use. Remote sensing is the process of deriving information through systems not in direct contact with the objects or phenomena of interest: image processing describes the manipulation of this raw data yielded by remote sensing systems. The combination of these technologies can yield a variety of spatial data.

There are two main providers of primary remote sensing data. The Earth Observation Satellite Observation Company (EOSAT) is the operator of the Landsat remote sensing satellite system and the distributors of the primary data (and some interpreted data) from this system. The information from this system is read with a pixel size of 30m x 30m. The French company Systéme Probatoire de la Observation de la Terre (SPOT) has a satellite in orbit and a distribution office in the United States. SPOT data are available with a 20m x 20m pixel size for color data and a 10m x 10m black and white pixel size. SPOT is also the only supplier of global Digital Terrain Modeling information. Both companies offer data for a variety of sizes of databases and a variety of media (digital/tape, paper, or film/transparency). The cost for digital system corrected map sheet information is $0.49 per square km for EOSAT data and $0.68 per square km for SPOT data.

The National Oceanic and Atmospheric Administration (NOAA) has used the Nimbus series of satellites to gather Advanced Very High Resolution Radiometer (AVHRR) data since 1978. AVHRR data are only sensed at 1.1 kilometer resolution, but offer the advantage of semidiurnal readings. The private vendor handling public access to AVHRR data notes that this information is not archived and contains encoded meteorological data; no land cover/land use information is available through AVHRR.

Remote sensing data can directly provide information on land use and land cover. Digital processing can enhance the satellite data to interpret this physical evidence to describe physical and cultural processes.

ATTACHMENT B - SAMPLE WIND ROSES

(Prepared by OAQPS/MRB)

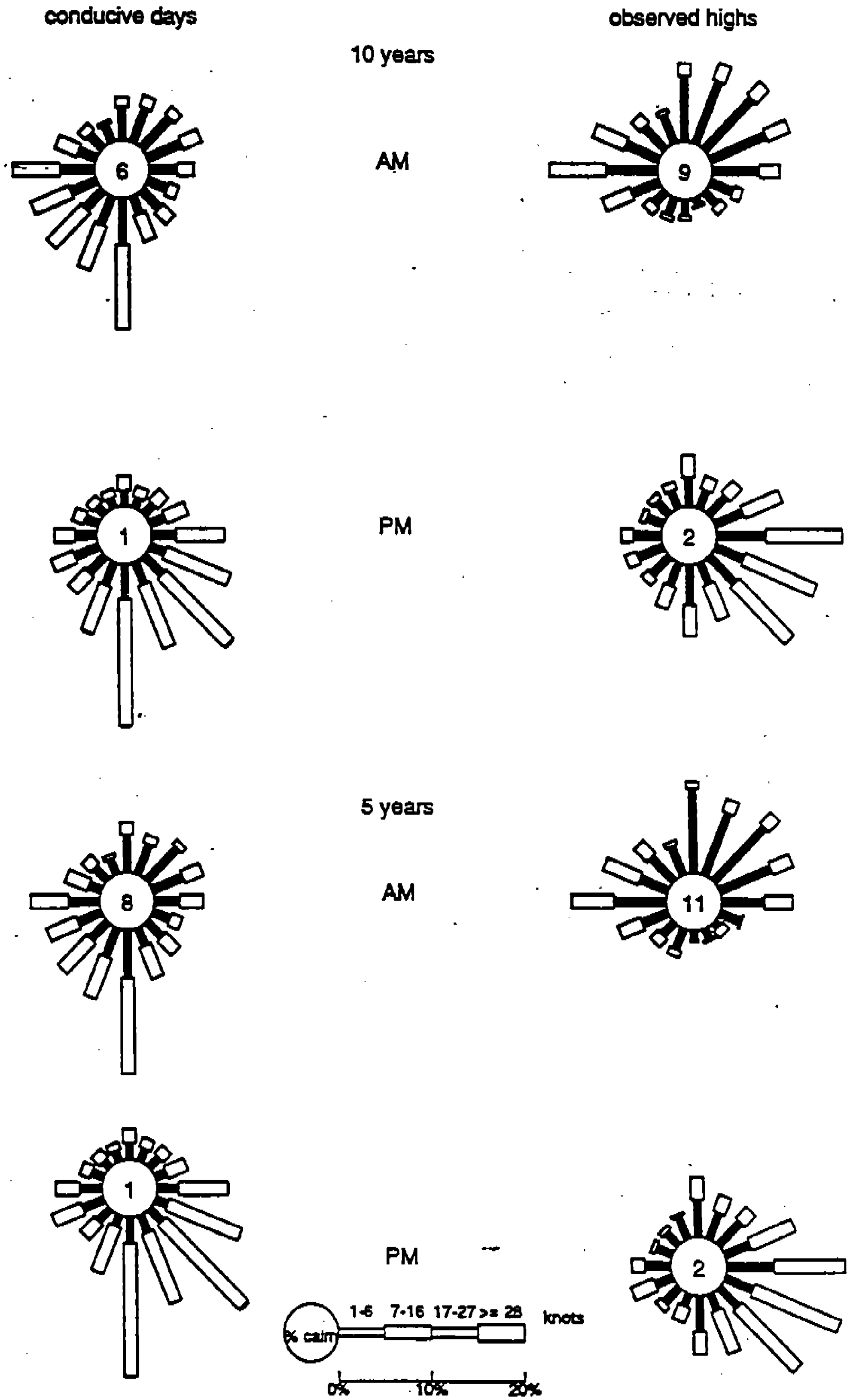


Figure B-1. Sample Wind Roses

ATTACHMENT C - GUIDANCE FOR SUBMITTAL OF EMISSIONS INFORMATION

- Introduction
- Point Source Emissions
- Area Source Emissions
- Mobile Source Emissions

ATTACHMENT C
PAMS MONITOR SITING - USE OF EMISSIONS RELATED DATA AND SURROGATES

1.0 INTRODUCTION

Presentation of a summary of the available SIP base year (1990) VOC and NO_x emissions by the predominant major source categories will provide an initial indication of the relative magnitudes of the source categories for which geographic distribution may be relevant to the selection of PAMS Site #2. Appropriate data for this highly aggregated level of summary should exist in the documentation of current SIP inventory submittals for the nonattainment area of concern and should be available directly from State personnel responsible for inventory preparation.

The primary purpose of obtaining and presenting emissions data summaries is to provide perspective on the relative distribution of the area's emissions across source types. This information will make it possible to develop priorities as to the sources for which geographical distributions are important in the process of locating PAMS Site #2. At a minimum, VOC and NO_x emissions totals should be presented for the following categories: (1) point sources, broken down into combustion and non-combustion sources; (2) area sources, broken down into population-related and industrial source types; and (3) mobile sources, with separate totals for highway and non-road sources (see Table C-1).

For most purposes related to PAMS station siting, the ideal form of geographic emissions information is a set of emissions density displays at a small grid scale, such as that which can be produced by emissions preprocessors for urban scale photochemical models. Where these types of presentations of emissions data are not available or where the resolution provided is not adequate for specific monitor siting considerations, alternatives are discussed below.

TABLE C-1. EXAMPLE AREA-WIDE EMISSIONS SUMMARY

	Ozone Season Weekday Emissions (tons/day) VOC		NO$_x$	
	tons/day	percentage	tons/day	percentage
POINT SOURCE				
Combustion	1	< 0.1	78	22.4
Non-combustion	9	1.5	1	< 0.1
AREA SOURCE				
Population-related	115	19.1	16	4.6
Industry-related	160	26.6	14	4.0
MOBILE SOURCE				
Highway vehicles	195	32.4	231	66.4
Non-road	22	3.7	8	2.3
BIOGENICS	100	16.6		
TOTALS	602	100*	348	100*

* Due to rounding percentages do not total 100.

2.0 POINT SOURCE EMISSIONS

Site #2 monitoring sites should be selected in areas that will allow accurate assessment of representative amounts of point source emissions (*i.e.*, emissions from facilities that generate 10 tons per year (tpy) VOC emissions or greater, or from facilities that generate 100 tpy or greater NO_x emissions. Smaller sources should be included if such data are available. A map identifying point emissions sources and monitoring sites is the best way to concisely present the required data. Maps should be generated at two levels of resolution: (1) a *macro* level, showing the entire SIP and monitoring area; and (2) a *micro* level, representing a neighborhood-level view of the monitoring sites and nearby point sources. VOC emissions are generated from many point source mechanisms (*e.g.*, combustion, evaporation, waste disposal) and therefore are generated at many emissions points. In comparison, NO_x is primarily a combustion byproduct, generated at a smaller number of emissions points. This fundamental difference in emission generation, coupled with the tendency of NO_x emission sources to mask VOC emissions, results in the need for separate maps identifying VOC and NO_x point emission sources.

2.1 MACRO-LEVEL MAPPING

Two macro-level maps should be submitted supporting Site #2 monitoring sites. One map will summarize the monitoring locations and VOC emissions levels, the other will relate monitoring sites to NO_x emission points.

2.1.1 VOC Map

Point source emissions should be presented on a map that indicates VOC emissions on a zip code level (see Figure C-1). In areas where the geographic areas represented by zip codes are too small to be clearly represented on a macro-level map, the sources could be identified by latitude/longitude. VOC emissions in each zip code may be identified by cross-hatching, color, or gray scale, and should be resolved into quartiles or higher resolution. VOC emissions for each facility in the SIP area should be available from EPA's Aerometric Information Retrieval System (AIRS) Facility Subsystem (AFS). Zip codes are typically included for AFS facilities[1]; in absence of zip codes, an approximation may be made using the latitude/longitude or Universal Transverse Mercator (UTM) data typical in AFS. Wind roses should be presented on the same page, in the same orientation, as the emissions map. Including wind roses will make evaluating potential site locations easier and more consistent. Software and climatological data for generating wind roses are available in the Support Center for Regulatory Air Models (SCRAM) on EPA's Technology Transfer Network bulletin board service (TTN).

[1]AFS includes zip codes for both *street* and *mailing* addresses. Care should be taken to use *street* addresses. This avoids inaccuracies caused when a facility submits the mailing address of a remotely located home office or when a rural facility maintains an in-city post office box.

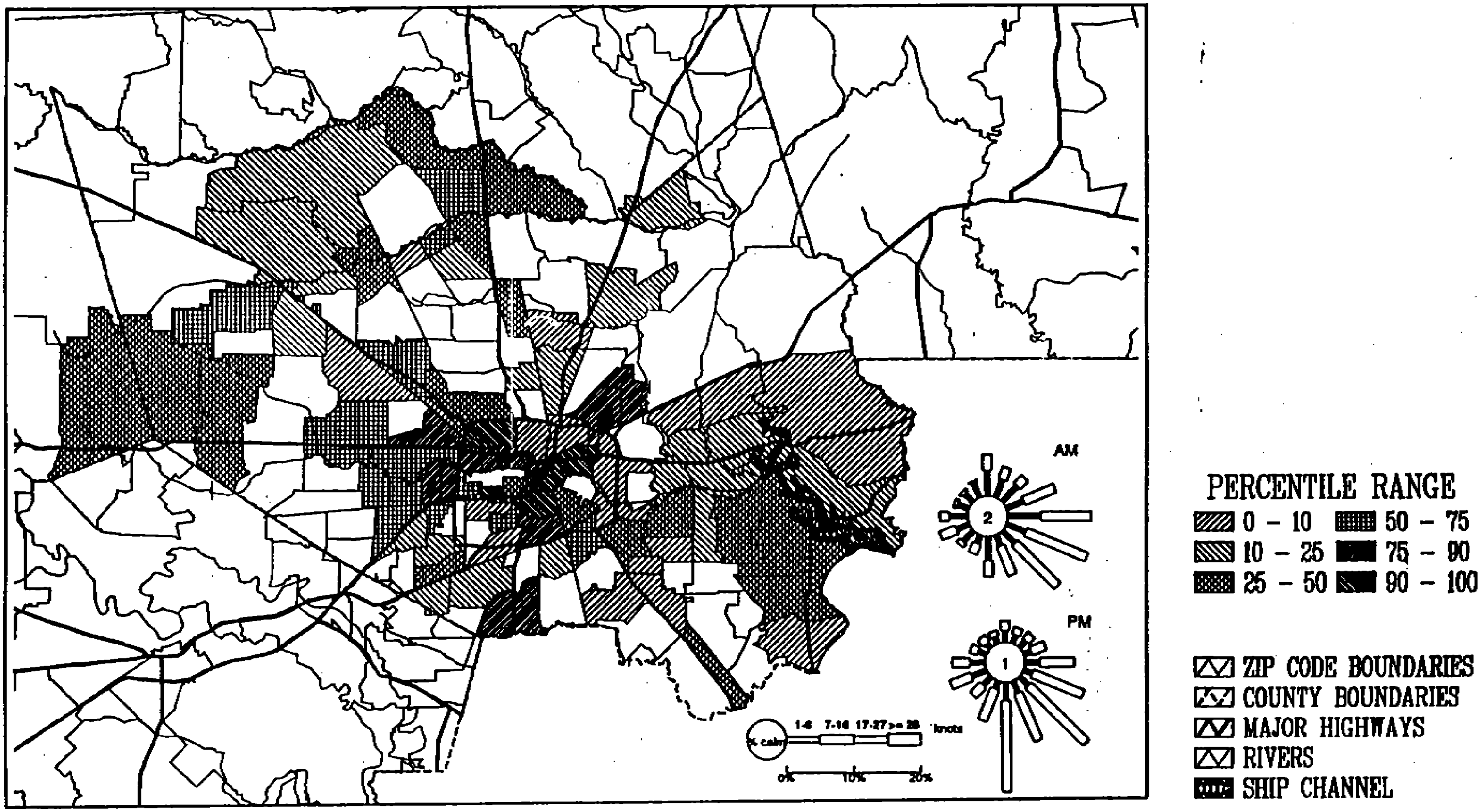


Figure C-1 VOC emission densities by ZIP code

2.1.2 NO_x Map

As stated earlier in this section, NO_x is a combustion byproduct and therefore is produced in a number of sources. Because the NO_x resulting from a source may react with, and therefore mask, ambient VOC levels, NO_x emissions points should be located and identified individually rather than aggregated to a zip code level so that the VOC measurements are not unduly influenced by a single source. NO_x emissions may be identified from AFS; facility location is typically reported using latitude/longitude or UTM data. The relative emissions from each source may be reported on the map by indicating the location of each facility with an indicator (circle, dot or spike) sized in proportion to its NO_x emissions. As in the case of the VOC maps, wind roses should be presented along with the maps.

2.2 MICRO-LEVEL MAPS

Micro-level maps show, on a neighborhood level, the location of the monitoring sites. Micro-level maps generally show a 4-5 km radius around the monitoring station (see Figure C-2). One micro-level map should be submitted for each monitoring location to be implemented within the next ozone monitoring season.

All point sources in the neighborhood of the monitoring site should be indicated on this map. As suggested for the NO_x macro map, the relative emissions from each source may be reported on the map by indicating the location of each facility with an indicator (circle, dot or spike) sized in proportion to its NO_x emissions. Color or shape may be used to indicate pollutant (VOC or NO_x). AFS may be consulted to determine emissions, pollutant and location (using plant address, latitude/longitude, or UTM code).

Wind roses should be printed on the same sheet as the map, oriented consistently with the map to facilitate evaluation.

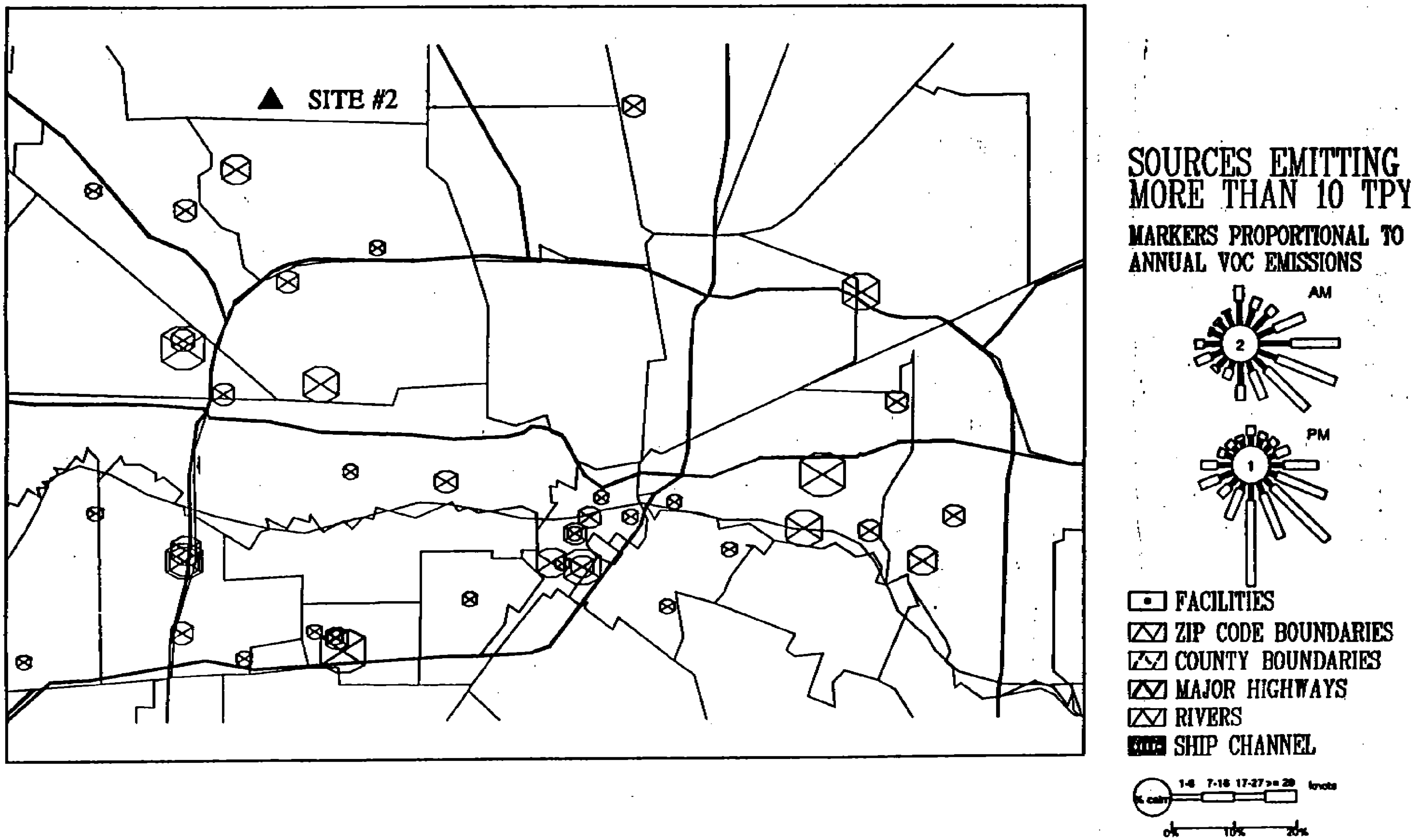


Figure C-2 Point source VOC emission intensities

3.0 AREA SOURCE EMISSIONS

Many types of emission sources are too small or too prevalent to be inventoried as point sources. Area source classifications allow estimation of emissions from such sources by using emission factors based upon substituting data descriptive of activity throughout a neighborhood or county for facility-specific data. While indicators of the distribution of area sources include varied data such as population, housing, land cover, land use, and sales of relevant commodities or products, surrogates for the area sources responsible for VOC emissions are most closely paralleled with population distribution. Therefore, maps representing the population distribution of a nonattainment region will indicate neighborhoods with the highest area source emission potential.

A work of caution is given, however, because the use of population as a surrogate activity level indicator is true for activities such as dry cleaning, architectural surface coating, small degreasing operations, and solvent evaporation from household and commercial products. However, per capita factors should not be used for sources whose emissions do not correlate well with population. In some instances, it might be better to use employment rather than population as a surrogate activity level indicator. Such factors are usually appropriate to estimate emissions for source categories which have an SIC code and for which employment data at the local level are available. Because large fractions of VOC emissions are covered by point source procedures, the emissions-per-employee factor is a secondary procedure to cover emissions from sources that are below the point source cutoff level.

There are several public sources of population and demographic data. The U.S. Department of Commerce, Bureau of the Census provides population data on paper, magnetic tape/disk, and optical disk. The data may be resolved to a variety of levels down to the size of the typical city block. Data at the census tract level, a region containing approximately 2,500 to 8,000 people, is sufficient for most area source estimates. Local and regional planning departments also maintain population data. Though the source of this data may also be the federal census, it may be aggregated into regions more relevant to the development and urbanization patterns of the metropolitan area.

4.0 MOBILE SOURCE EMISSIONS

Mobile sources consist of on-road sources (often referred to as "highway vehicles") and off-road sources, which include aircraft, railroads, vessels, and "non-road" sources such as farm and construction equipment, utility engines, lawnmowers, etc. While on-road mobile sources are expected to be a dominant source category in all PAMS areas, the importance of off-road sources may vary.

4.1 ON-ROAD MOBILE SOURCES

Potential sources of information on the geographic distribution of on-road mobile source emissions include:

- Highly resolved grid-based on-road mobile source emissions data such as that which would be prepared for photochemical oxidant model input (gridded emission densities) (see Figure C-3)
- Emissions estimates by road system links and traffic generation and destination areas, resulting from the use of a calibrated travel demand model in combination with the EPA MOBILE on-road mobile source emission factor model
- Link- and area-based transportation activity estimates (vehicle miles travelled, trip starts and trip ends) from a calibrated travel demand model
- Traffic count maps indicating the level of traffic on the area's road system. (see Figure C-4)

Traffic flow maps may be available from the local organization(s) charged with performing traffic counts. Some of these maps use bandwidths to indicate the relative level of traffic on individual roads, while others feature the transportation grid system with actual traffic counts written in numerically at the counted locations. The level of detail of these maps will vary with the intensity of the local counting program, but typically includes only the larger roads in the area (such as major arterials and expressways).

4.2 OFF-ROAD MOBILE SOURCES

The level of effort and detail appropriate for development of geographic distributions for off-road mobile sources is a function of the magnitude of these emissions in the area of concern, which would be shown by emissions summaries. The highly disparate nature of the individual off-road source types complicates geographical allocation of this category. If the off-road category represents a critical or significant portion of the overall emissions of a given pollutant, and locations of these emissions have not already been characterized for air quality modeling, investment of some effort in geographical allocation for PAMS purposes may be justified. The basic steps would include (1) obtaining a detailed listing of emissions from individual off-road source types, (2) assigning priorities based on their emission levels, (3) developing reasonable surrogates or logical locational characteristics for each source type, and (4) graphically depicting the distribution of emissions represented by these surrogates or characteristics.

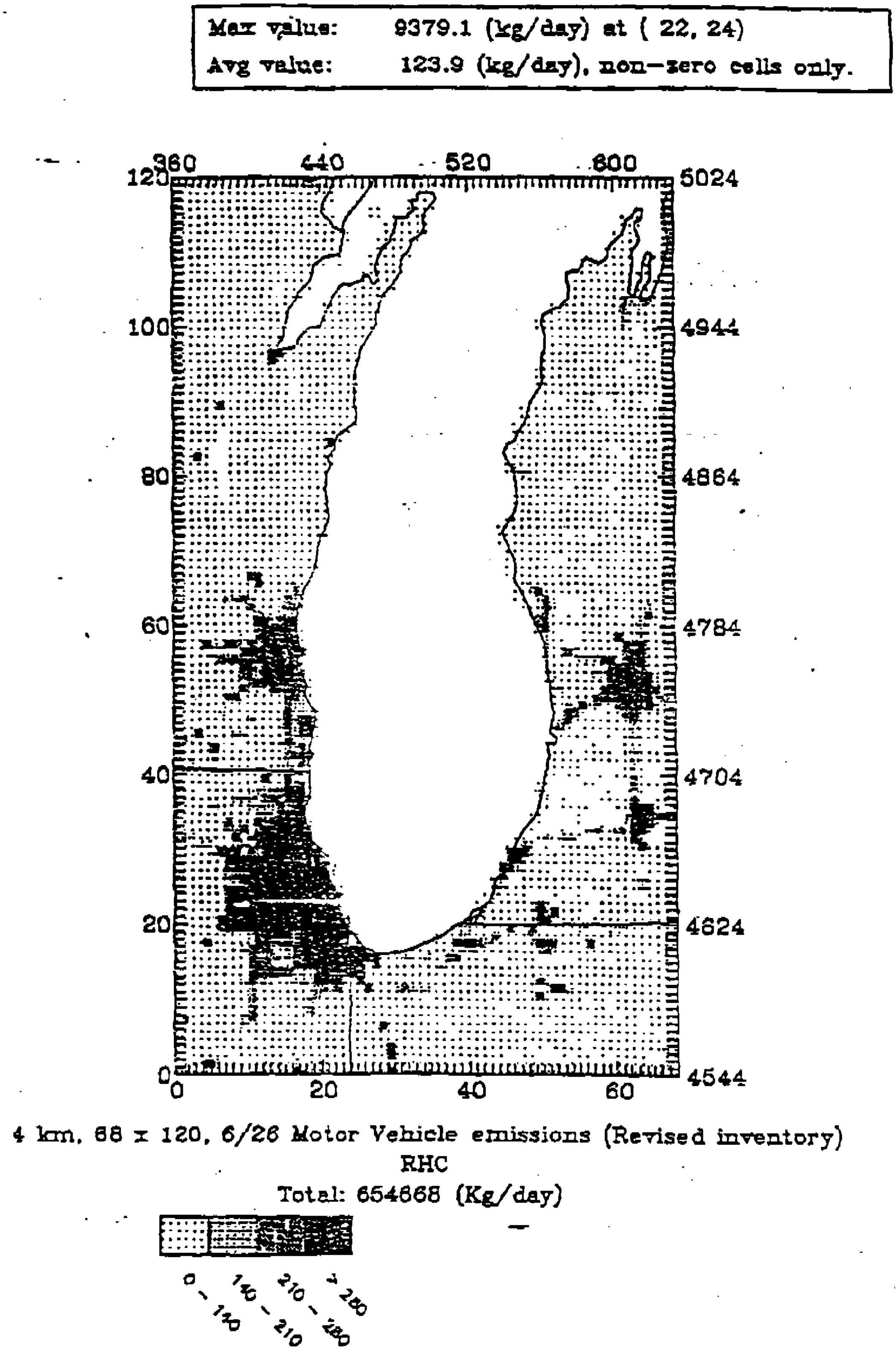


Figure C-3 Grid-based mobile emissions data map

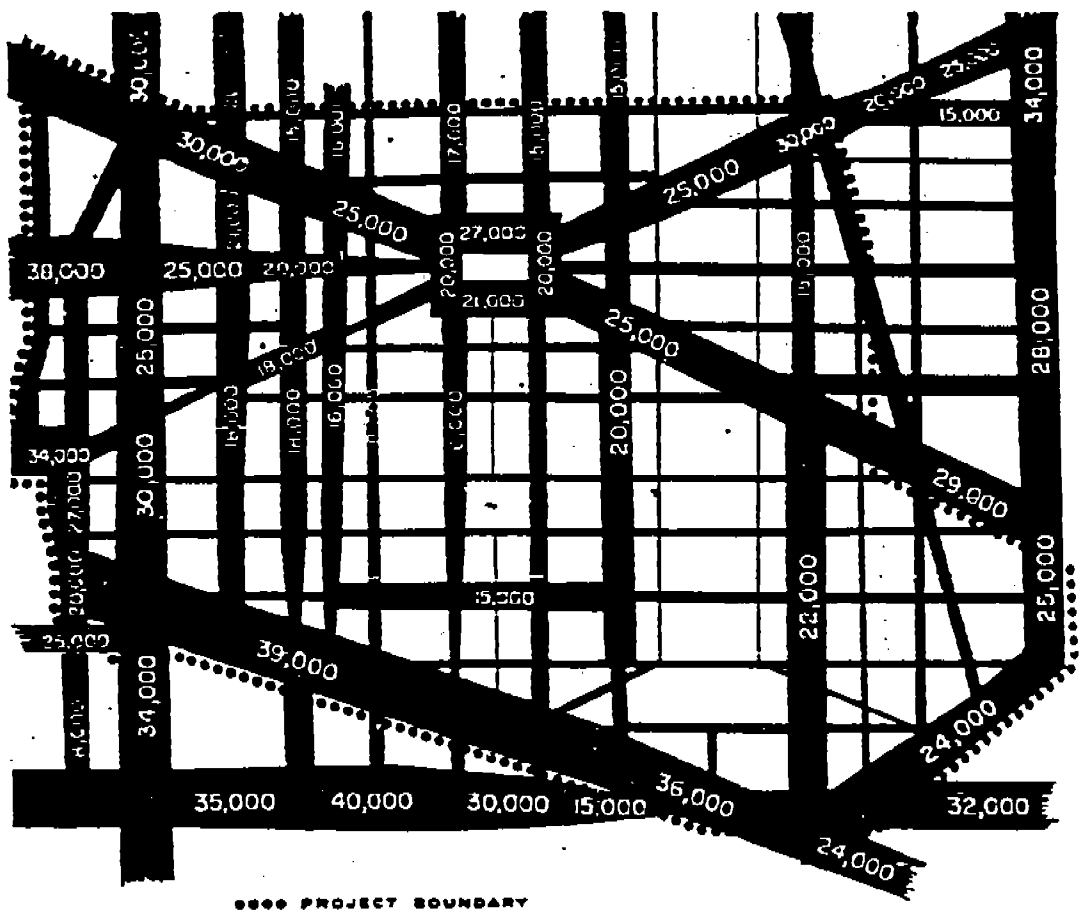


Figure C-4 Sample traffic density map for a downtown area

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
ATMOSPHERIC RESEARCH AND EXPOSURE ASSESSMENT LABORATORY
RESEARCH TRIANGLE PARK
NORTH CAROLINA 27711

September 15, 1993

MEMORANDUM

SUBJECT: Requirements and Guidance for PAMS Meteorological Station in New Brunswick, New Jersey

FROM: Jerry H. Crescenti, Physical Scientist
Human Exposure Modeling Branch (MD-56)

Jerry Crescenti

TO: N. Ogden Gerald
Office of Air Quality and Planning Standards (MD-14)

This memo is in response to your request for information about the requirements and incorporation of a 30 meter meteorological tower and upper air profilers into the Photochemical Assessment Monitoring System (PAMS) which will be located at Rutgers University in New Brunswick, New Jersey. There are two categories of data needed to fulfill PAMS meteorological requirements - surface and upper air.

The variables which are required from a ground-based system include wind speed, wind direction, air temperature, relative humidity, barometric pressure and incident solar radiation. Most of the guidance summarized below is taken from the Quality Assurance Handbook for Air Pollution Measurement Systems, Volume IV: Meteorological Measurements (1983, revised 1989, EPA-600/4-82-060) and the On-Site Meteorological Program Guidance for Regulatory Modeling Applications (1987, EPA-450/4-87-013).

The primary objective of instrument siting (horizontal and vertical probe placement) and exposure (spacing from obstructions) is to place the sensor in a location where it can make precise measurements that are representative of the general state of the atmosphere in that region under study. The choice of a site must be made with a complete understanding of the regional geography, the sources being investigated, and the potential uses of the data being collected. Ideally, the 30 meter tower should be located in an open level area. In terrain with significant topographic features, different levels of the tower may be under the influence of different meteorological regimes at the same time. If this is the case, such conditions should be well documented.

The standard exposure of a wind sensor over level, open terrain is 10 meters above the ground. Open terrain is defined as an area where the horizontal distance between the instrument and any obstruction is at least ten times the height of that obstruction. An obstruction may be man-made (e.g., building) or natural (e.g., trees). The wind sensor should be mounted on a mast at the top of the tower or on a boom projecting horizontally out from the tower. If the sensor is mounted on a boom, then it should be located at a horizontal distance at a minimum of twice the maximum diameter or diagonal of the tower away from the nearest point on the tower. The boom should project in the direction which provides the least distortion for the predominant wind direction. For example, the boom should be aligned in a northwesterly or southeasterly direction if the predominant wind is from the southwest.

Air temperature and humidity sensors should be mounted over a plot of open level ground at least 9 meters in diameter. The ground surface should be covered with non-irrigated or unwatered short grass or, in areas which lack a vegetation cover, natural earth. The surface must not be concrete, asphalt or oil-soaked. If there is a large paved area nearby, these sensors should be at least 30 meters away from it. These sensors should be located at a distance from any obstructions of at least four times their height. Areas of standing water should also be avoided as well as steep slopes, ridges or hollows. The standard heights are 2 and 10 meters, but additional levels may frequently be required in air quality studies. These sensors must be housed in well ventilated solar radiation shields (forced aspiration is preferable) at a horizontal distance from the nearest point on the tower of at least the diameter of the tower.

As for the barometric pressure sensor, there is no particular siting guidance available since the data is seldom utilized in EPA studies. However, the barometric pressure is needed when computing variables such as specific humidity, potential temperature and air density. I would suggest that the sensor be place indoors with the data acquisition system. One end of a rubber tube should be attached to the sensor's pressure port and the other ended vented to the outside of the trailer or shelter so that pressurization due to the air conditioning or heating system is avoided.

Solar radiation measurements should be taken in a location free from any obstruction which can either cast shadows or reflect sunlight on to sensor. This means that there should be no object above the horizontal plane of the sensing element that could possibly cast a shadow on it (including the tower). In addition, the pyranometer should not be placed near light colored walls or artificial sources of radiation. Usually, a tall platform or a roof make suitable locations for sensor placement. However, since it is not desirable to have a large obstruction such as a building in the vicinity of the tower, then the best strategy is to place the pyranometer directly south of the tower and its guy wires. There is no height requirement for this sensor.

The variables required for upper air monitoring to support PAMS include profiles of horizontal wind velocity, vertical wind velocity, and air temperature. Also needed is an estimate of the mixing layer height and stability class of the atmospheric boundary layer. However, temporal and spatial density of these variables have not been clearly defined. Rather, the amount of data to be acquired is a function of the field study objectives and numerical model input requirements.

There are two types of wind profiling systems. The first type is a RADAR which transmits a 915 MHz electromagnetic signal and has a range of approximately 90 to 3000 meters. The second type is a SODAR which transmits a 2 to 5 KHz acoustic signal and has a range of about 60 to 600 meters. Both systems transmit their respective signals in pulses. Each pulse is both reflected and absorbed by the atmosphere as it moves upwards. The height range of each pulse is determined by how high it can go before the signal becomes so weak that the energy reflected back to the antenna can no longer be detected. That is, as long as the reflected pulses can be discerned from background noise, meaningful wind velocities can be obtained. The attenuation of the pulses are functions of signal type, signal power, and atmospheric conditions. A Radio Acoustic Sounding System (RASS) utilizes a combination of electromagnetic and acoustic pulses to derive an air temperature profile in the range of about 90 to 1200 meters.

For upper air monitoring, remote sensing is quickly becoming the method of choice. However, while these profiling systems have been approved and used to develop meteorological databases required as input for dispersion models, there is a distinct void in terms of guidance needed to help potential users and the regulatory community. Because of their unique nature and constant evolution, the available EPA guidance for SODARs is more generic than that which already exists for many well established in-situ meteorological sensors. Almost no guidance is currently available on RADAR and RASS systems.

Since SODARs utilize sound transmission and reception to determine the overlying wind field, a clear return signal with a sharply defined atmospheric peak frequency is required. Thus, consideration of background noise may put limitations on where a SODAR can be located. External noise sources can be classified as active or passive, and as broad-band (random frequency) or narrow-band (fixed frequency). General background noise is considered active and is broad-band. If loud enough, it can cause the SODAR software to reject data because it can not find a peak or because the signal-to-noise ratio is too low. The net effect is to lower the effective sampling rate due to the loss of many transmission pulses. Radian (assuming that this company is providing the profiling systems) should be consulted as to what noise level would be acceptable for the SODAR. A qualitative survey should be conducted to identify any potential noise sources. A quantitative noise survey may be necessary to determine if noise levels are within the instrument's minimum requirements.

Examples of active, broad-band noise sources include highways, industrial facilities, power plants, and heavy machinery. Some of these noise sources have a pronounced diurnal, weekly or even seasonal pattern. A noise survey should at least cover diurnal and weekly patterns. Examination of land-use patterns and other sources of information may be necessary to determine if any seasonal activities may present problems.

Examples of active, fixed-frequency noise sources include rotating fans, a back-up beeper on a piece of heavy equipment, birds and insects. If these noise sources have a frequency component in the SODAR operating range, they may be misinterpreted as good data by the SODAR. Some of these sources can be identified during the site selection process. One approach to reducing the problem of fixed frequency noise sources is to use a coded pulse, i.e., the transmit pulse has more than one peak frequency. A return pulse would not be identified as data unless peak frequencies were found in the return signal the same distance apart as the transmit frequencies. Radian can provide information on whether or not the SODAR is capable of such a task.

Passive noise sources are objects either on or above the ground (e.g., tall towers, power transmission lines, buildings, trees) that can reflect a transmitted pulse back to the SODAR antenna. While most of the acoustic energy is focused in a narrow beam, side-lobes do exist and are a particular concern when antenna enclosures have degraded substantially. Side-lobes reflecting off stationary objects and returning at the same frequency as the transmit pulse may be interpreted by the SODAR as a valid atmospheric return with a speed of zero. It is not possible to predict precisely which objects may be a problem. Anything in the same general direction in which the antenna is pointing and is higher than 5 to 10 meters may be a potential reflector. It is therefore important to construct an "obstacle vista diagram" prior to SODAR installation that identifies the direction and height of potential reflectors in relation to the SODAR. This diagram can be used after some data have been collected to assess whether or not reflections are of concern at some SODAR height ranges. Note that reflections from an object at distance X from an antenna will show up at height $X\cos(\alpha)$, where α is the tilt angle of the antenna from the vertical.

The RADAR, SODAR and RASS antennas should be aligned and tilted carefully as small errors in orientation or tilt angle can produce unwanted biases in the data. True North should also be established for antenna alignment. Installation of the antennas should not be permanent since problems are very likely to arise in siting the profilers in relation to the tower and other objects that may be in the area. One final consideration is the effect of the instrument on its surroundings. The sound pulse from a SODAR and RASS is quite audible and could become a nuisance to residents who might happen to live near the installation site.

A joint effort is currently underway between EPA's Atmospheric Research and Exposure Assessment Laboratory in Research Triangle Park, North Carolina and NOAA's Wave Propagation Laboratory in Boulder, Colorado to further develop QA and QC guidance for wind and temperature profiling systems. Once established, this information will provide standard and consistent operating procedures which will lead to the acquisition of high quality, interpretable, and scientifically defensible data sets.

cc: Alan H. Huber
William F. Hunt
Robert G. Kellam
Dale A. Pahl
Larry J. Purdue
Brian D. Templeman

APPENDIX J

AIR QUALITY INDICATORS

United States
Environmental Protection
Agency

Office of Air Quality
Planning and Standards
Research Triangle Park, NC 27711

EPA-450/4-81-015
February 1981

Air

EPA

U.S. ENVIRONMENTAL PROTECTION AGENCY INTRA-AGENCY TASK FORCE REPORT ON AIR QUALITY INDICATORS

EPA-450/4-81-015
February 1981

U.S. Environmental Protection Agency Intra-Agency Task Force Report on Air Quality Indicators

by

W.F. Hunt, Jr. (chairman), G. Akland, W. Cox, T. Curran,
N. Frank, S. Goranson, P. Ross, H. Sauls, and J. Suggs

U.S. Environmental Protection Agency
Office of Air, Noise, and Radiation
Office of Research and Development
Office of Planning and Management
Region 5

Prepared for

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air, Noise, and Radiation
Office of Air Quality Planning and Standards
Research Triangle Park, NC 27711

February 1981

This report is issued by the Environmental Protection Agency to report technical data of interest to a limited number of readers. Copies are available - in limited quantities - from the Library Services Office (MD-35), U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711; or, for a fee, from the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.

Publication No. EPA-450/4-81-015

CONTENTS

CONTENTS (continued)

FIGURES

FIGURES (continued)

FIGURES (continued)

TABLES

1. OVERVIEW

The Intra-Agency Task Force on Air Quality Indicators was established to recommend standardized air quality indicators and statistical methodologies for presenting air quality status and trends in national publications. As a first step, the members of the task force identified topics of concern and prepared a series of issue papers on these topics; these papers discuss the background and current status of each issue, develop recommendations, and identify areas that need additional work. These individual papers make up the remaining sections of this document.

To put the activities of the task force in perspective, it should be noted that on May 10, 1979, EPA promulgated regulations for ambient air quality monitoring and data reporting. These regulations were a result of the groundwork of the Standing Air Monitoring Work Group (SAMWG), and reflect EPA's concerns about data quality, timeliness, and consistency from one area to another. Specific provisions in these regulations instituted the routine reporting of precision and accuracy information to aid in characterizing data quality, the designation of specific sites in state and local agency monitoring networks to be used in national trends analyses, and the increased standardization of siting criteria to ensure greater uniformity. All ambient air quality data received by EPA should reflect these changes by 1981; the data bases currently used in EPA's analyses are in transition. In a sense, the monitoring community has identified and begun to implement improvements in the air quality data bases, and now those responsible for analyzing the data must ensure that the best use is made of these improvements.

1.1 RECOMMENDATIONS

In developing recommendations, two common concerns were apparent. The first involved data bases that do not yet exist (e.g., precision and accuracy information). Since it is premature to recommend how these data should be used in reporting air quality, the Task Force has simply identified the group that should take the initiative in developing methods for using the information. The second concern involved the relative merits of standardization. The data from a given air quality study can be analyzed by a wide variety of statistical techniques. In many cases, different approaches are equally acceptable. In fact, in certain cases, there are statistical techniques that are recommended but seldom applied. The important point here is that an emphasis on standardization should not discourage the development and application of new techniques. Consequently, these recommendations should be viewed as a set of general principles rather than a set of inflexible rules. If a particular approach satisfies the intent of these recommendations, it is satisfactory; if it does not, an explanation should be included in the analysis to say why alternative techniques were used.

The recommendations are grouped in four categories: data base, data analysis, data interpretation, and data presentation. In each category, both general and specific points are presented.

1.1.1 Data Base

In general, each analysis should indicate what data were used. For small studies, specific sites can be named; for large studies, it will suffice to indicate the data source (e.g., the National Air Data Branch, NADB) and the selection criteria used to choose sites. Specific recommendations are listed below.

1. Precision and Accuracy (Section 2) - EPA does not require the submission of precision and accuracy information until 1981; therefore no guidance on its use will be given at this time. It is recommended that ORD/EMSL take the lead in developing procedures for using this information.

2. Data Screening (Section 3) - Statistical procedures for detecting outliers are available, and some have been incorporated into SAROAD. Under the new monitoring regulations and management plan for the National Air Monitoring Stations (NAMS), users will eventually be able to assume that NAMS data quality has been verified. In the interim, however, the user should apply appropriate screening procedures to the data for any small-scale analysis; for large-scale analyses, the user may rely on robust statistical techniques that will minimize the potential impact of anomalous data.

3. Site Selection (Sections 4 & 5) - The NAMS will provide a usable, quality assured, standardized data base for trends--particularly national trends. Composite values of NAMS data will provide a

useful index for national trends assessment. In the interim, the user must select sites on the basis of specific criteria which ensure adequate completeness and seasonal balance. The criteria should be stated clearly in the analysis.

1.1.2 Data Analysis

The analysis should be structured so that results are stated in terms of statistical significance. Analyses that have no statistical basis should indicate that they do not and why they do not. Those analyses with a statistical basis should indicate the statistical approach used. More specific recommendations concerning data analysis follow.

1. Choice of Summary Statistics (Section 6) - Summary statistics should reflect the appropriate air quality standard and not be biased due to sample size. If an analysis requires the use of a statistic that is biased with sample size, care must be taken to ensure that comparisons over time or across sites are not affected by differences in sample sizes.

2. Comparability of the Data Base (Sections 5 & 6) - Any trends analysis should be structured so that results are not attributable to the data base varying with time. Interpolated data may be used for the visual presentation of trends, but trend statistics should be based on actual data unless the effect of interpolation can be quantified.

3. Trend Techniques (Section 6) - Standard statistical techniques such as Chi-square, nonparametric regression, aligned-rank tests, analysis of variance, and time series are all acceptable means of assigning probability statements to trends analyses. The primary concern is that the tests used are statistical in nature, not which tests are used. However, EPA groups need to take a more active role in applying various techniques to air data to assess the relative merits of different procedures.

1.1.3 Data Interpretation

An analyst can facilitate the interpretation of air quality monitoring data from the existing NAMS network by using other sources of information that help explain why an air quality trend has or has not taken place or why there are significant differences between sets of air quality data. To better assess the effectiveness of EPA's emission control program, the agency should collect data on all variables that impact air quality in at least two major urban areas.

1.1.4 Data Presentations

Data presentations should be consistent with the analysis, and should be adequately labeled so that they can stand alone.

1. Choice of Scales - Distortion of scales should be avoided. In general, the pollutant concentration axis should start at zero concentration.

2. Distinguish No-Data Cases - Presentations involving shading (e.g., maps) should clearly indicate cases with no data as a distinct category.

1.2 FUTURE WORK

In view of the transition occurring in the air quality data bases, it is recommended that this task force be continued during 1981, when NAMS data and precision and accuracy data will be received initially by EPA. During this time, the following tasks should be performed:

1. Assessment of Statistical Manpower - To ensure that the task force recommendations can actually be implemented and are not merely idealized goals, it will be necessary to appraise the available and planned technical resources. A breakdown of resources needed should be provided, with particular attention to the in-house resources needed to provide continuity and technical guidance.

Lead Group: OPM

Target Date: March 81

2. Use of Precision and Accuracy Information - The eventual use of precision and accuracy information needs to be better defined. Although these data are not currently being received by EPA, similar preliminary information is available to EMSL. These data should be examined, and a plan should be developed on how this type of information can be incorporated into EPA's use of air quality data. Consideration should be given to the feasibility of eventually establishing national performance standards for precision and accuracy data.

Lead Group: ORD/EMSL

Target Date: July 81

3. Use of Site Information in Trends Analyses - An important feature of the NAMS data base is the detailed information available describing individual sites. To routinely make efficient use of this information, it will be necessary to identify the relevant site parameters and to develop computer software to link the site information with the air quality information. Attention should be given to stratifying the data into broad classifications needed for more refined data analysis.

Lead Group: OAQPS/MDAD

Target Date: September 81

4. <u>Assessment of Statistical Software</u> - Several statistical software packages are available that could be used for air quality data analysis. To ensure that EPA statistical manpower is efficiently utilized an assessment should be made of what statistical software is applicable. In particular, this assessment should determine: (1) what statistical packages are being used and to what extent and (2) if the best statistical packages are being employed and if not, why not.

Lead Group: ORD/EMSL/RTP

Target Date: September 81

5. <u>Presentation of Data</u> - Because the Task Force is recommending specific types of data presentations a pilot study should be initiated to ensure that these recommendations are feasible to implement on a routine basis. Attention should be given to identifying computer programs that would facilitate these presentations and if gaps exist to develop the necessary programs to the extent possible with existing resources.

Lead Group: Region V

Target Date: September 81

2. KNOWN MEASURES OF DATA UNCERTAINTY (Precision and Accuracy)

Concepts of precision and accuracy have been the subject matter for many presentations and publications. ASTM Committee E-11 on quality control of materials discussed definitions and implications of these two ideas over a 10-year period.[1] Complete agreement on the meanings of precision and accuracy is unlikely to be found in the literature on scientific measurement systems. In assessments of the quality of ambient air data, EPA uses estimates of precision to characterize the relative capability of the monitoring system to repeat its results when measuring the same thing and estimates of accuracy to characterize the closeness of an observation to the "truth."[2]

Beginning January 1, 1981, EPA will require all state and local reporting agencies to calculate precision and accuracy estimates in a prescribed manner and to qualify all data entered into the EPA data bank with quarterly precision and accuracy estimates. Some, but not all, of the components necessary for estimating precision and accuracy (Appendix A, 40-CFR-58)[3] are available. Currently, collocated samplers and single concentration precision checks provide data which can be used to estimate the precision of manual and automated samplers, while audits of flow rates and laboratory analytical measurements provide data which can be used for estimating accuracy. Perhaps improved methods for estimating these parameters will be developed as a result of the increased attention to assessing air quality data.

Table 1 presents the types and frequencies of special checks for precision and accuracy by pollutant measurement method; Table 2 displays the concentration range for each audit level.[3]

2.1 NEED FOR QUALIFYING AIR POLLUTION DATA

Over many years, the air pollution data bank has grown into a gigantic body of computerized records of concentrations of pollutants measured at sites across the Nation at points in time. Necessarily, great amounts of attention and expense have been devoted to devising systems to process the data into a

TABLE 1. SPECIAL CHECKS AND AUDITS FOR ESTIMATION OF PRECISION AND ACCURACY

	Precision	Accuracy (local audit)*	
Automated analyzers (SO_2, CO, NO_2, O_3)			
Type check	One concentration	Three or four concentrations	
Frequency	Biweekly	25% of the analyzers each quarter; at least one per quarter	
Scope	All monitoring instruments	All analyzers each year	
Manual methods		Type of audit	
		Flow	Analytical
Type check SO_2 NO_2 TSP	Collocated samplers at two sites	NA NA One level	Three levels Three levels NA
Frequency	Each monitoring day	25% of the sites each quarter; at least once per quarter	Each analysis-day; at least twice per quarter
Scope	Two sites (of high concentration)	All sites each year	NA

*See Table 2 for audit levels.

TABLE 2. CONCENTRATION RANGES FOR AUTOMATED ANALYZER AUDITS

	Concentration range, ppm	
Audit level	SO_2, NO_2, O_3	CO
1	0.03-0.08	3-8
2	0.15-0.20	15-20
3	0.40-0.45	40-45
4	0.80-0.90	80-90

computerized retrievable form. Also, high priority has been given to the development and refinement of a system for collecting data from state and local agencies and sending the data through EPA Regional Offices to the NADB computer system. Inevitably, air pollution control administrators and affected industries are concerned about the quality of the data that result from this great expenditure of resources, especially pollutant measurements that may represent exceedances of National Ambient Air Quality Standards (NAAQS) or that have other serious implications.

As a result of continued programs of quality assurance and technological improvements in monitoring and analysis, today's ambient air data are doubtlessly more representative of true concentrations than those of some years ago. However, these improvements in data quality cannot be quantified because no routine, standardized data assessment program has been in effect. Implementation of the program described above should provide the means to evaluate progress in measuring and recording ambient air data and should give managers a higher level of confidence in making decisions based on air pollutant measurements.

2.2 SCHEDULE FOR IMPLEMENTING THE PRECISION AND ACCURACY PROGRAM

All designated NAMS sites are to begin operation January 1, 1981 (Appendixes A and E, 40-CFR-58). On July 1, 1981, the first quarterly report is due into EMSL; the first quarterly ~~summary~~ report from EMSL to the EPA Regional Offices is due September 1, 1981; and the first annual report is scheduled for July 1, 1982. The remaining State and Local Air Monitoring Stations (SLAMS) are to be phased into the Precision and Accuracy Reporting System as soon as possible after January 1, 1981. The distinction of SLAMS and NAMS is of little relevance here, since precision and accuracy data relate only to agencies.

2.3 USES OF PRECISION AND ACCURACY DATA

Confidence in conclusions concerning air quality trends will be more scientifically defensible with the precision and accuracy data because of the increased capability to test for statistical significance in trends analyses. Also, additional interpretive insights may be gained; for example, if accuracy does not change significantly, the quality assurance program may be ruled out as a factor affecting trends.

For many trends study purposes, the analyst must set an arbitrary limit (e.g., ±15%) on sites to be included in an analysis. The analyst must make judgments in light of the purpose and the user requirements. In all cases, the analyst should test to determine any effect of precision and accuracy information on conclusions.

It is premature at this time to attempt to develop definitive criteria for using precision and accuracy data in determining trends, nonattainment status, and so forth. Inclusion of precision and accuracy data in analyses may appear to add to the overall levels of uncertainty, but experience over time is needed to establish criteria for setting limits on the data to be used in a particular situation.

Precision and accuracy data will permit comparisons of data quality within and between monitoring networks and within and between States and Regions. Also, precision and accuracy data together with data from EPA's Ambient Air Performance Audit Program should provide the means for effective evaluations of analytical laboratories hired by State and local agencies.

A great deal of interest centers on the possible uses of precision and accuracy data to report probability intervals about peak and mean estimates of air quality data. If the peak value is not an outlier, there is no problem with the probability interval. Long-term (3 years) studies may be required to verify the assumptions that the sample of precision and accuracy information is representative and that extreme values are within the population. There should be no problem in reporting probability intervals for mean values. In any case, the assessed quality of air pollution measurements should be included in all relevant reports and publications.

2.4 REFERENCES

1. Precision Measurement and Calibration, Statistical Concepts and Procedures. National Bureau of Standards Special Publication 300. Volume 1. February 1969.

2. Rhodes, R. C. Precision and Accuracy Data for State and Local Air Monitoring Networks: Meaning and Usefulness. Paper presented at the 73rd APCA Annual Meeting, Montreal, Canada, June 1980.

3. Appendix A. Quality Assurance Requirements for State and Local Air Monitoring Stations (SLAMS). FR Vol. 44, No. 92, pp. 24574-27581, May 10, 1979.

3. DETECTING AND REMOVING OUTLIERS

An outlier is an observation that does not conform to the pattern established by other observations. This pattern may be a scatter plot, frequency histogram, time series, or simple listing. The parent population from which the observations are drawn is usually assumed to behave according to a certain probabilistic model, and the observations are usually assumed to be drawn at random or at least independently. The agreement between the observations and the parent population depends both on the correctness of the underlying assumptions and on certain aspects governing the selection process. Outliers should not be isolated from other problems of statistical analysis; they should be included among anomalies such as nonadditivity, nonconstant variance, bias, temporal drift, or wrong model specification. A correction in another problem can often solve the outlier problem.

3.1 CAUSES OF OUTLIERS

Three basic ways an outlier can occur are (1) mistakes in readings, (2) wrong model specification, and (3) rare deviation.[1]

Mistakes in readings can occur during any stage of data processing, but the more common mistakes (e.g., transcription errors) usually occur early in the processing; during data coding or key punching, transcription errors often go unnoticed. Unusual readings from instruments may be caused by power failure or surges, improper calibration, breakdowns, torn filters, contamination, chemical interaction, leaks, and so forth. Not adhering to an experimental plan or design can affect recorded data. Mistakes occur both in totally automated data gathering processes and in those that rely on human consideration or intervention.

With the exception of glaring mistakes that have obvious explanations, model specification usually is the basis for deciding if a discordant observation is an outlier. In fact, most tests currently being used routinely by EPA to detect outliers are actually testing some type of nonnormality.[2] For example, the null hypothesis of all the tests listed in Table 3, with the exception of the gap test, is that the population from which the observations are being drawn is specified by a normal distribution model. References 3 and 4 describe additional tests which specify normal, lognormal, or Weibull models. Detection of an outlier by any one of these tests is paramount to declaring the parent population has a different distribution than was originally assumed. Of course, the final conclusion would be that the observation is an outlier.

In any parent population, some observable events have a low probability of occurring. These events are usually associated with the extremes (tails) of the distribution. Although the probability of obtaining one of these observations is small, it is possible for such an event to occur. Being a rare occurrence, the observation is almost always treated as an outlier.

3.2 STATISTICAL PROCEDURES FOR IDENTIFYING OUTLIERS

Statistical procedures for identifying outliers are used for different reasons. One reason is to justify what would have been done anyway (i.e., to reject observations that an experienced investigator would intuitively reject). Another reason for using statistical tests is to provide algorithms to the computer[2,5] for scanning large sets of data that would be impractical to scan by visual inspection. At present, 5122 SLAMS, 1269 NAMS, and 68 Inhalable Particulate Network (IPN) sites are producing data which require validation. In addition, special studies such as the Philadelphia IP Study, the National Forest Ozone study, and the Across Freeway Study (LACS) are collecting large quantities of continuous and 24-hour data. Whether the data are collected routinely or for special studies, a procedure for detecting outliers is necessary to strengthen the validity of conclusions reached during data analysis.

Table 3 lists several statistical tests for objectively screening a set of data and identifying possible outliers. These tests are currently being applied to routine data storage systems within EPA. These tests, except for the studentized range, the studentized t-test, and the Shewhart control chart require no external or independent parameter estimates; conclusions are based solely on the data at hand. Except for the Gap Test,[2] these tests for outliers are tests for normality. All tests are aimed at high values relative to some measure of spread based on the sample. Where routine screening of data is necessary, a battery of several tests is advisable.

The Dixon ratio test[6] was the first of many standard statistical procedures which have been found to work well in air data screening. Some work better than others, depending on sampling frequencies and durations. For example, the Shewhart control chart was found superior to the Dixon ratio in screening 24-hour measurements.[2] These and other screening procedures constitute the Air Data Screening System (ADSS), which is now being implemented in 27 States through the Air Quality Data Handling System.[2]

In another screening program, several statistical procedures are currently being applied to SAROAD data.[2] With the exception of the gap test, these tests have one thing in common--they all assume that the observations are samples from a single normal population with specific location and shape parameters. Tables and charts are used to identify values that have a low probability of occurring if all observations were taken from the same population; these values are flagged for further investigation.

To demonstrate how outliers can be identified, several tests, in Table 3 are applied to the following set of mass data in the 0-15 micron size range gathered over a 5-month period in Birmingham, Alabama, using a dichotomous sampler.

24-hour samples		Monthly values	
Date	$\mu g/m^3$	Average ($\bar{x}$)	Range (R)
08/01/79	43.8		
08/07/79	66.7		
08/25/79	17.8		
08/31/79	45.3	43.4	48.9
09/06/79	92.6		
09/12/79	34.8		
09/30/79	38.4	55.3	57.8
10/06/79	47.5		
10/12/79	64.7		
10/18/79	36.2		
10/24/79	30.6		
10/30/79	16.9	39.2	47.8
11/05/79	36.6		
11/23/79	15.3	26.0	21.3
12/05/79	101.8		
12/11/79	35.0	68.4	66.8

Dixon Ratio[2,3]

Calculated: $R_{22} = \frac{X_n - X_{n-2}}{X_n - X_3} = \frac{101.8 - 66.7}{101.8 - 17.8} = 0.418.$

Tabulated: $r_{0.90}$ (i.e., $\alpha = 0.10$) = 0.454 for n = 16.

Thus: $R_{22} < r_{0.90}$.

Shewhart Control Chart[2,3]

Range: $LCL_R = D_3\bar{R} = (0)(43.95) = 0.$

$UCL_R = D_4\bar{R} = (2.57)(53.95) = 112.95.$

In this example, the largest integer less than the average was used (n = 3).

Average: $LCL_{\bar{X}} = \bar{\bar{X}} - A_2\bar{R} = 40.95 - (1.02)(43.95) < 0$ (use $LCL_{\bar{X}} = 0$ in this case).

$UCL_{\bar{X}} = \bar{\bar{X}} + A_2\bar{R} = 40.95 + (1.02)(43.95) = 85.78.$

Thus the monthly average (68.4) and range (66.8) for December are not considered outliers.

Chauvenet's Criterion[9]

$\bar{X} = 45.25$

$S = 25.00$

$$C = \frac{X_n - \bar{X}}{S} = \frac{101.8 - 45.25}{25.00} = 2.26.$$

The critical value of C for sample size n = 16 is 2.154 with an α-level of 0.016 for a one-tailed test, thus 101.8 is an outlier. Chauvenet's criterion is not recommended for samples of n < 7 for a two-tailed test and for samples of n < 4 for a one-tailed test because it tends to flag too many valid observations.

Grubbs Test[7,8]

Calculated: $SS = 9372.26$, $SS_{16} = 5961.16$, and $SS_{15,16} = 3161.25$.

Thus: $L_1 = \frac{SS_{16}}{SS} = 0.64$ and $L_2 = \frac{SS_{15,16}}{SS} = 0.34$

Tabulated: $L_1 = 0.576$ and $L_2 = 0.405$ for n = 16.

Therefore, at the $\alpha = 0.05$ significance level, only the highest value (101.8) can be rejected as an outlier.

Coefficient of Skewness[10]

Calculated: $$B = \frac{n \sum_{i=1}^{n} (X_i - \bar{X})^3}{\left[\sum_{i=1}^{n} (X_i - \bar{X})^2\right]^{3/2}} = 1.0$$

Normally, this test for skewness is not used for n small (say < 25); however, the value of B exceeds the tabulated value 0.71 for n = 25 (the tabulated value for n = 16 will be < 0.71) and therefore, normality is rejected at the α = 0.05 significance level.

After excising 101.8 and 92.6, normality can no longer be rejected; both values are outliers based on this test.

Studentized T-Test[9]

The standard deviation (S) required for testing studentized deviates was estimated from the August, September, and October data to be 21.52. November and December data are grouped together because there are only two observations in each month. The suspected outlier is 101.8 $\mu g/m^3$.

$$T = \frac{X_n - \bar{X}}{S} = \frac{101.8 - 47.18}{21.52} = 2.54.$$

The critical value ($\alpha = 0.05$) for examining one outlier among four samples, relative to the prior S based on 11 degrees of freedom, is $t_{0.05}(4,11) = 2.24$. Thus 101.8 $\mu g/m^3$ is an outlier.

Studentized Range[9]

The S required for the studentized range test was estimated from the same set of data as the studentized t-test. November and December data are grouped together as before, and the suspected outlier is again 101.8 $\mu g/m^3$.

Calculated: $W = X_n - X_1 = 101.8 - 15.3 = 86.5.$

Tabulated: $W = q_{0.05}(4,11)S = 4.26(21.52) = 91.68$

Thus: 101.8 is not an outlier.

Some tests are more suited to hourly than to 24-hour data. Table 3 summarizes the recommended applications of the commonly used tests. For examples of how other tests are applied, consult the references.

References 3 and 4 describe nonparametric tests for comparing two or more data sets to aid in identifying a data set that does not conform to the pattern established by the other data sets. Reference 4 also discusses the use of nonparametric tests for a single sample. However, identifying an outlier by a nonparametric test is inconsistent with our definition of an outlier; an outlier is an atypical observation because it does not conform to a model which we hypothesize to describe the data. If we have no model in mind, it is difficult to describe what we mean by an outlier. Hence, nonparametric techniques are recommended only for the comparison of data sets.

3.3 RECOMMENDED TREATMENT OF OUTLIERS

No observation should be rejected solely on the basis of statistical tests, since there is always a predictable risk (i.e., the α-level) of rejecting "perfectly good" data. Compromises and tradeoffs are sometimes necessary, especially in routinely scanning large amounts of data; in these situations, no statistical rule can substitute for the knowledge and judgement of an experienced analyst who is thoroughly familiar with the measurement process.

In any data-collecting activity, all data must be recorded along with any notes that may aid in statistical analysis. Gross mistakes should be corrected if possible before performing calculations in the final analysis; if a mistake cannot be explained or corrected, it is not always wise to discard the reading as though it had never occurred. Further experimental work may be needed, since a gross error in the observations can bias the analysis in all but the most robust statistical procedures. In addition, data collected under differing conditions should not be combined for identifying outliers unless the experiment was designed (e.g., factorial design) to handle such data with an appropriate model during final analysis.

In performing the calculations in the final statistical analysis, the question of what weight, if any, to assign a discordant value is difficult to answer in general terms. An acceptable explanation for an outlier should preclude any further use of the value. Sometimes it is obvious that an observation does not belong even though there is no explanation for its existence. Although the value should not be used in further calculations, it should be mentioned in the final report.[7]

Experienced investigators differ greatly on these matters.[9] Excluding "good" data may not be as serious as including "bad" data and then excluding any questionable observations in further calculations at the risk of losing information on estimates and introducing some bias. Reducing sample variation (increasing precision) may be preferred over introducing a slight bias, especially if the bias is theoretically estimable. Using robust air quality indicators that are not strongly influenced by outliers may avoid many argumentive, subjective decisions.

3.4 CONCLUSIONS

The data analyst should not assume that all outliers represent erroneous values. Some outliers occur because the analyst has used the wrong probabilistic model to characterize the data. For example, outlier tests based on assumptions of normality may be inappropriate for nonnormal data sets. In addition, there is always a finite chance that an extreme value will occur naturally. The analyst should carefully investigate these possibilities before discarding outliers. The use of robust air quality indicators is recommended since they decrease the need for detecting outliers. Reference 4 discusses several outlier tests, their performance, and provides a brief summary of comparative information for each test, including pertinent references.

3.5 REFERENCES

1. International Encyclopedia of Statistics. Vol. 2. The Free Press, New York, 1978. pp. 1039-1043.

2. Guideline Series. Screening Procedures for Ambient Air Quality Data. EPA-450/2-78-037, July 1978.

3. Nelson, A. C., D. W. Armentrout, and T. R. Johnson. Validation of Air Monitoring Data. EPA-450/2-78-037, August 1980.

4. Barnett, Vic and Toby Lewis. Outliers in Statistical Data. John Wiley and Sons, Inc. New York, New York, 1978.

5. Northrop Services, Inc. A Data Validation Program for SAROAD. Cont. No. 68-02-2566. ESG-TN-78-09, December 1978.

6. Dixon, W. J. Processing Data for Outliers. Biometrics. 9(1):74-89, March 1953.

7. Grubbs, F. E., and G. Beck. Extension of Samples, Sizes and Percentage Points for Significance Tests of Outlying Observations. Technometrics 14(4):847-854, November 1972.

8. Tietjen, G. L., and R. H. Moore. Some Grubbs-Type Statistics for the Detection of Several Outliers. Technometrics 14(3):583-597, 1972.

9. Natrella, M. G. Experimental Statistics. Chap. 17. National Bureau of Standards Handbook 91, 1963.

10. Duncan, A. J. Quality Control and Industrial Statistics. Richard D. Irwin, Inc., 1965.

11. Quality Assurance Handbook for Air Pollution Measurement Systems. Volume I. EPA-600/9-76-005. U.S. Environmental Protection Agency, Environmental Monitoring and Systems Laboratory, Research Triangle Park, N.C., January 1976.

TABLE 3. OUTLIER TESTS

Test name	Test criterion	Definition of terms	Basis for test	Critical values	Recommended application
Dixon ratio[3,9,11]	$r_{10} = (X_n - X_{n-1})/(X_n - X_1)$, $3 < n < 7$ $r_{11} = (X_n - X_{n-1})/(X_n - X_2)$, $8 < n < 10$ $r_{21} = (X_n - X_{n-2})/(X_n - X_2)$, $11 < n < 13$ $r_{22} = (X_n - X_{n-2})/(X_n - X_3)$, $14 \leq n \leq 25$	X_n = maximum value of n samples, X_{n-1} = second maximum value, X_1 = minimum value, other subscripts indicate order from lowest to highest	A value too large compared to other values in the sample; underlying distribution is normal	Special tables for r_{ij}[9,11]	24-h samples on a monthly basis; values may be individuals or averages of individuals
Shewhart control chart[3,9,10,11]	Range : $LCL_R = D_3\bar{R}$, $UCL_R = D_4\bar{R}$ Average: $LCL_{\bar{X}} = \bar{\bar{X}} - A_2\bar{R}$, $UCL_{\bar{X}} = \bar{\bar{X}} + A_2\bar{R}$	$\bar{R}$ = average range of several groups of samples, $\bar{\bar{X}}$ = overall average, LCL = lower control limit, UCL = upper control limit, A_2, D_3, D_4 are constants used to calculate 3σ control limits for $\bar{X}$ and R, depending on sample size	A value outside control limits based on average and range of previous samples taken in a time sequence and assuming an underlying normal distribution	Special tables[10,11] for D_3, D_4, and A_2 to calculate 3σ control limits depending on sample size	Monthly averages and ranges of 24-h samples
Chauvenet's criterion[9]	$C = \dfrac{(X_n - \bar{X})}{S}$	X_n = largest value in a set of n samples, S = standard deviation, i.e., $S = \left[\dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}\right]^{1/2}$ $\bar{X}$ = average of n samples, $\bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$	A value too large compared to the sample average, underlying distribution is normal and that a mistake occurs on the average once in 2n samples	Special tables with α-levels for two-tailed tests given as 1/2n, and 1/4n for 1-tailed tests	1-h values on a daily basis; not recommended for n<4 for a 1-tailed test

(continued)

TABLE 3 (continued)

Test name	Test criterion	Definition of terms	Basis for test	Critical values	Recommended application
Coefficient of skewness[19]	$B = \frac{n \sum_{i=1}^{n} (X_i - \bar{X})^3}{\left[\sum_{i=1}^{n} (X_i - \bar{X})^2\right]^{3/2}}$,	X_j = value of the jth sample, n = total number of samples	"Tail effect" of an unknown number of observations; underlying distribution is normal	Special tables[19] for 0.05 & 0.01 points of the distribution of B assuming normality	1-h values on a daily basis; may be applied sequentially to several outliers
Studentized t-test[9]	$T_n = \frac{(X_n - \bar{X})}{S}$	X_n = largest value in a group samples, $\bar{X}$ = average, S = standard deviation, estimated from prior or concurrent samples (i.e. independent of the numerator	A value too large compared to the mean, relative to the standard deviation; underlying normal distribution	Special tables[9] of percentage points of critical values $t_q(n,v)$	1-h values on a daily basis
Studentized range[9]	$W = X_n - X_1$	X_n = maximum value in a set of n samples, X_1 = minimum value, W = range of a set of n samples	Extreme values too far apart; underlying distribution is normal	Special tables[9] for relative range distribution percentage points; an independent estimate of the standard deviation is required	1-h values on a daily basis
Sequential difference[9]	1-h: $D_i = X_i - X_{i-1}$ 24-h: $D_i = X_i - X_{i-24}$	X_i = ith sample where i refers to the hour, X_{i-1} = sample taken in the hour preceding i, X_{i-24} = sample taken on the previous day corresponding to the ith hour of the present day	A value in a time series being significantly larger than the previous hour's sample under assumption of normality. A value in a time series being larger than the previous day's sample during the same hour under assumption of normality	3σ limits based on previous data & assumption of normality	1-h values 24-h values

(continued)'

TABLE 3 (continued)

Test name	Test criterion	Definition of terms	Basis for test	Critical values	Recommended application
Grubbs test[7,8]	$L_2 = \frac{SS_{n-1,n}}{SS}$	$SS_{n-1,n}$ = sum of squares corrected for the avg of the samples where the two highest values have been excluded. SS_n = sum of squares corrected for the avg and excluding the highest value. SS = sum of squares corrected for the avg excluding no values	One or more values too large compared to the avg; normal distribution	Special tables[7,8] for L_1 and L_2 percentage points; F-tables may be used if mean squares are used in place of sums of squares	1-h values on a daily basis; multiple outliers in a univariate case; better than sequential application of other tests in preventing the "masking effect"; NOTE: Tietjen & Moore[8] expand tables to cover more than two outliers at a time
Gap test[2]	$p = e^{-n\lambda k}$	k = length of a gap in a frequency distribution to the right of which n values lie. λ = parameter of the exponential distribution based on the 50th and 95th percentile of each data set. p = probability that n values will lie to the right of a gap of length k	One or more values lying to the right of a gap too large compared to the other observations; assuming independent samples from an exponential distribution	Special software; identifies gaps with probabilities of less than 0.01 of having values greater than the gap	1-h values for any length of time; multiple outliers in a univariate case; NOTE: Another gap test not currently being routinely used was developed by Tietjen & Moore[8]; this test is based on Grubbs type statistics where the largest gap to the right of the mean is first identified & the number of observations above the gap are tested

4. AREA OF COVERAGE AND REPRESENTATIVENESS

4.1 BACKGROUND

Over the decade of the 1970's the number of ambient air pollutant monitors increased dramatically from approximately 1800 in 1970 to approximately 8000 in 1979.[1] The lack of uniform criteria for station locations, probe siting, sampling methodology, quality assurance practices, and data handling procedures resulted in data of unknown quality. In preparing "national" air quality trends data bases for the major pollutants, the number of monitoring sites changed constantly, reflecting the growth in State monitoring networks during this period. A conglomeration of different types of sites--rural, industrial, residential, commercial, etc--evolved with no national plan. Some urban areas of the country had extensive monitoring while others did not.

In October 1975, at the request of the Deputy Administrator of EPA, a Standing Air Monitoring Work Group (SAMWG) was established. The Work Group was to critically review and evaluate current air monitoring activities and to develop air monitoring strategies which would be more cost effective, would help to correct identified problems, would improve overall current operations, and would adequately meet projected air monitoring goals. Members of the Work Group represent State and local air pollution control agencies and EPA program and regional offices.

SAMWG's review indicated that the current ambient monitoring program is basically effective in providing information for support of State Implementation Plan (SIP) activities. Several areas were identified where deficiencies existed, however. The principal areas where corrections are needed are summarized below.

1. Lack of uniformity in station location and probe siting, sampling methodology, quality assurance practices, and data handling procedures have resulted in data of unknown quality.

2. Existing regulations coupled with resource constraints do not allow State and local agencies sufficient flexibility to conduct special purpose monitoring studies.

3. Resource constraints and the diversity of data needs frequently result in untimely or incomplete reporting of air quality data. Adequate air quality data for national problem assessments and routine trend analyses are in some cases not available to agency headquarters until 12-18 months after each calendar quarter.

4. In some cases, data are being reported to the EPA central data bank from more stations than are absolutely necessary for adequate assessment of national control programs and analysis of pollutant trends.

The recommendations of SAMWG later became the basis for the Federal monitoring regulations promulgated on May 10, 1979.[2] These regulations require EPA to:

- set stringent requirements for a refined national monitoring network in areas with high population and pollutant concentrations to provide a sound data base for assessing national trends;
- give the States flexibility to use resources freed from SIP monitoring work to meet their own needs;
- establish uniform criteria for siting, quality assurance, equivalent analytical methodology, sampling intervals, and instrument selection to assure consistent data reporting among the States;
- establish a standard national pollutant reporting index and require its use for major metropolitan areas; and
- require the submission of precision and accuracy estimates with air quality data to enable better interpretation of data quality.

These regulations should produce a streamlined, high-quality, more cost-effective national air monitoring program.

The States are required to establish a network of stations to monitor pollutants for which National Ambient Air Quality Standards (NAAQS) have been established. Each network is to be designed so that stations are located in all areas where the State and the EPA Regional Office decide that monitoring is necessary. The stations in the network are termed State and Local Air Monitoring Stations (SLAMS).

Data summaries from the network are to be reported annually to EPA. Data from a subset of SLAMS to be designated as National Air Monitoring Stations (NAMS) are to be reported quarterly to EPA.

4.2 NETWORK DESCRIPTION

4.2.1 The SLAMS Network

The SLAMS network should be designed to meet a minimum of four objectives:

1. To determine the highest concentrations expected in each area covered by the network;
2. To determine representative concentrations in areas of high population density;
3. To determine the impact of ambient pollution levels from significant sources or source categories; and
4. To determine background concentrations.

Each monitoring site is required to be identified by location and type of surroundings as well as by monitoring objective and spatial scale of representativeness. The spatial scale of representativeness is described in terms of the physical dimensions of the air parcel sampled by the monitoring station throughout which actual pollutant concentrations are reasonably similar; the scale adjectives are micro, middle, neighborhood, urban, regional, national, and global.

4.2.2 The NAMS Network

The NAMS stations are selected from the SLAMS network to emphasize urban and multisource areas. The primary objective for NAMS is to monitor areas where pollutant levels and population exposure are expected to be highest, consistent with the averaging time of the NAAQS. Accordingly, NAMS fall into two categories:

1. Stations in area(s) of expected maximum concentrations; and
2. Stations with poor air quality and high population density but not necessarily in area(s) of expected maximum concentrations.

For each urban area where NAMS are required, both categories of stations must be established. If only one NAMS is needed to monitor suspended particulates (TSP) and sulfur dioxide (SO_2), the first category must be used. The NAMS are expected to provide superior data for national policy analyses, for trends, and for reporting to the public on major metropolitan areas. Only continuous instruments will be used at NAMS to monitor gaseous pollutants.

Siting requirements and definitions vary with pollutant, but NAMS are required only in urban areas with populations of at least 50,000 where pollutant concentrations are known to exceed the secondary NAAQS. The number of urban areas monitored will vary with pollutant, as indicated below.

Pollutant	Urban areas	NAMS monitors
Total suspended particulates (TSP)	212	636
Sulfur dioxide (SO_2)	160	244
Carbon monoxide (CO)	77	121
Ozone (O_3)	85	208
Nitrogen dioxide (NO_2)	33	65

The NAMS network is being selected on the basis of experience with past monitoring data and primarily attempts to measure the highest pollutant levels associated with each urban area. The word "associated" indicates that, with transport-effected pollutants such as O_3 or TSP, the highest levels may occur in areas of low population densities. The location selection should be periodically reviewed, based on historical monitoring data, meteorological conditions, and changes in emission patterns. For example, there is definite evidence that the highest O_3 levels in the Los Angeles basin have been moving upwind in the last few years, probably because of the changing composition of tailpipe pollutants. Changing fuel usage and new construction may have also contributed to this shift.

4.3 REPRESENTATIVENESS

The question has often been asked as to what constitutes a "national" trend? In previous National Air Quality, Monitoring and Emission Trends Reports,[3,4,5] all sites with data available in the National Aeromatic Data Bank (NADB), that could meet an historical data completeness criteria, were selected for the national trend. In the case of total suspended particulate (TSP) as many as 3000 sites could meet an historical completeness criteria. These sites come from a variety of networks representing urban areas, rural areas, and large point sources. Geographical coverage was largely weighted by population; that is, the more populated areas generally had more monitoring sites. In analyzing the national trend each site was weighted equally.

In contrast to TSP, the automotive related pollutants--carbon monoxide (CO) and ozone (O_3)--had less than 250 trend sites meeting historical completeness criteria as late as 1977.[5] Although California had a disproportionate number of these trend sites, each site was weighted equally in developing the national trend.

~~In~~ each of the three cases, the sites used for analyzing the national trend represented a mixture of siting criteria. Consequently, it was difficult to determine what the resulting trend really represented.

The major difficulty in selecting sites that define a "national" trend is determining what mix of sites represents national air quality. How should geographic distribution be determined--by population or by land mass? What types of sites--center city, suburban, residential, commercial, industrial, rural, remote, etc.--should be in the national trend and what proportions are appropriate?

4.3.1 The NAMS Solution

A partial solution to this dilemma lies in the NAMS network. Part of the problem with a "national" trend is a semantic one--how is it defined? The NAMS can resolve this problem because they can be defined. For each of the criteria pollutants, NAMS are located in either the areas of expected maximum concentration or the areas of high population density. Consequently, the NAMS lend themselves to stratification into these two site populations. Additional siting information is available through the NAMS management information system, which will allow for the use of covariate data (such as traffic counts in the case of CO) for the first time.

The NAMS are particularly appropriate for characterizing national trends for urban sites located in areas of expected maximum concentration or high population density. For example, the CO NAMS located in areas of expected maximum concentration make up a clearly defined population and could serve as a good indicator of the success (or failure) of the automotive emission control program.

4.3.2 The SLAMS and Detailed Urban Area Analyses

While the NAMS can be used to define national urban trends in areas of expected maximums or high population density, the SLAMS can be used for detailed urban area analyses. If one is trying to determine changes in air quality in an urban area, then an examination of both spatial and temporal change is in order. The SLAMS networks lend themselves to these types of analyses. EPA has published several guidelines useful in determining spatial and temporal trends; they should be consulted before initiating these types of analyses.[6,7]

4.4 REFERENCES

1. Monitoring and Air Quality Trends Report, 1974. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards. Research Triangle Park, N.C. Publication No. EPA-450/1-76-001. February 1976.

2. Federal Register, Vol. 44, February 8, 1979, pp. 8202-8237.

3. National Air Quality and Emission Trends Report, 1975. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards. Research Triangle Park, N.C. Publication No. EPA-450/1-76-002. November 1976.

4. National Air Quality and Emissions Trends Report, 1976. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C. Publication No. EPA-450/1-77-002.

5. National Air Quality, Monitoring and Emissions Trends Report, 1977. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C. Publication No. EPA-450/2-78-052. December 1978.

6. Guideline on Procedures for Constructing Air Pollution Isopleth Profiles and Population Exposure Analysis. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C. Publication No. EPA-450/2-77-024a. October 1977.

7. Users Manual for Preparation of Air Pollution Isopleth Profiles and Population Exposure Analysis. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C. Publication No. EPA-450/2-77-024b. October 1977.

5. DATA COMPLETENESS AND HISTORICAL CONTINUITY

Data completeness for summary statistics and historical continuity for trend data are two concerns of the data analyst. In both cases, the quantity of data available affects the uncertainty of the analytical results. In current practice, criteria are invoked in the data screening steps to select minimally acceptable data sets. This section explores the effect of missing data on the uncertainty of analytical results, and discusses criteria for ensuring adequate data completeness and historical continuity.

5.1 DATA COMPLETENESS

5.1.1 Background and Purpose

The number of air quality values produced by ambient monitors is usually fewer than the maximum number possible. Missing values result from the intermittent sampling schedules for manual methods, instrument failure, downtime for calibration, or human error. The temporal balance and the sample size of the resultant data set can seriously affect the validity of the sample and the uncertainty of its summary statistics.

Published criteria have been used by EPA to establish the validity of data sets for summarizing and analyzing air quality data. Such criteria should minimize the uncertainty associated with air quality summary statistics. Unfortunately, for each pollutant the same completeness criteria are used for all summary statistics, despite the fact that the uncertainty associated with a summary statistic varies with the type of statistic and its variance properties.

In a sense, validity criteria provide a capability for identifying data produced by poorly operating instruments and for screening data samples that may otherwise yield misleading or incorrect estimates of air quality levels. Data requirements defined by the criteria should involve the type of summary statistic (e.g., annual mean or maximum daily average) as well as its intended application (e.g., trends analysis or status assessment with respect to the standard). In general, fewer data are needed to determine an annual average than a short-term statistic, and fewer data per year are needed to estimate a long-term trend than a yearly status.

Currently, the NADB uses different validity criteria[1] for different time periods and sampling approaches, as shown in Table 4. The criteria for each sampling approach are intended to consider both the characteristics in the data collection and the primary objective of the monitoring; however, for specific applications there are inconsistencies in the criteria for different approaches. This section discusses only the criteria for summary statistics for periods of 3 to 12 months.

TABLE 4. NADB VALIDITY CRITERIA[1]

Continuous Sampling (1-h, 2-h, and 4-h data)
Quarterly statistics - 75% or 1642 hours
Annual statistics - 75% or 6570 hours
Intermittent Sampling (24-h data)
Quarterly statistics
Five samples per quarter
If 1 month has no values, at least two values in other months
Annual statistics - four valid quarters

5.1.2 Origin of NADB Criteria

Criteria for intermittent sampling were formulated on the basis of a bi-weekly sampling schedule which was used by the National Air Surveillance Network (NASN) up to 1972. Nehls and Akland recommended that more stringent criteria be applied when the sampling schedule is every 3rd or 6th day.[2]

For continuous sampling, the origin of the 75 percent criteria is not clear. Early Federal Air Quality Publications[3] used a 50 percent criterion to report summary statistics. Nehls and Akland suggested 75 percent completeness and a more structured basis for summary statistics, as shown in Table 5; these criteria require as few as 3402 hourly observations, or 39 percent of the possible hours.

TABLE 5. CONTINUOUS MEASUREMENT SUMMARY CRITERIA[2]

Time interval	Minimum requirement
3-h running average	3 consecutive hourly values
8-h running average	6 hourly values
24-h	18 hourly values
Monthly	21 daily averages
Quarterly	3 consecutive monthly averages
Yearly	9 monthly averages with at least 2 monthly averages per quarter

5.1.3 Characteristics of Air Quality Sampling

Intermittent 24-hour sampling generally follows a fixed systematic or pseudo-random schedule which is intended to provide representative coverage of a time period with only a fraction of the total possible observations. By design, data from intermittent sampling can provide good estimates of an annual mean, but may severely underestimate the peak values. Trying to improve the estimate of the peak observation by intensifying the sampling during periods of high pollution levels is known as episode monitoring. When these unscheduled episode data are combined with the scheduled data, bias can be introduced into summary statistics.

Early intermittent sampling for TSP and other pollutants by the NASN was biweekly on a modified random schedule which yielded 20 samples a year. Later the schedule was modified to ensure equal representation for each day of the week. After 1972, samples were collected every 12th day yielding a maximum of 30 to 31 samples a year. EPA's current recommended sampling schedule for TSP is once every 6 days. Most agencies seem to follow this schedule. Among 2882 sites with 1978 data that meet the NADB validity criteria, over 60 percent produced 40 to 60 samples, and less than 10 percent sample more often than 1 in 3 days.

Continuous hourly monitoring, by providing a more complete representation of air quality, should provide much better estimates of the true annual mean and short-term peak values. In reality, continuous monitoring is often incomplete, and can yield biased estimates of long-term behavior. Missing data often occur in blocks of consecutive days or weeks. Sometimes a pollutant such as O_3 is monitored during only part of the year; only seasonal statistics are appropriate in these cases.

TABLE 6. DATA COMPLETENESS FOR CONTINUOUS SO_2 MONITORING, 1973

Minimum quarterly completeness (%)	Number of sites: Annual completeness <75%	≥75%	Total
25	41	--	41
26-32	11	--	11
33-49	40	4	44
50-66	14	18	32
67-74	2	10	12
75-100		84	84
Total	108	116	224

As an example of actual monitoring performance, data completeness for 1973 SO_2 data is presented in Table 6. Of 224 sites, only 116 reported 75 percent of the total hourly observations. According to the percentage of completeness for 1973, 56 of the 108 sites not meeting the 75 percent annual completeness criterion had at least one-third of the total observations in each calendar quarter.

5.1.4 Characteristics of Air Quality Data

Air quality data are known to exhibit temporal components generally in the form of diurnal, weekly, or seasonal cycles. The variances for each component may be different, thus an arbitrarily selected sample of the entire series may not yield unbiased, minimum variance statistical estimates if the sample does not represent each portion of each time component equally.

For example if a diurnal pattern always shows the highest concentrations between 0800-1200 hours, but the sampling does not include any observations during these hours, simple sample statistics will have a negative bias. A sample requires temporal balance to be representative.

With a seasonal pattern of nonuniformly distributed data, the more extreme the seasonal variation, the larger the potential bias. In addition, the larger the sampling imbalance across the seasons, the larger the potential bias. If a seasonal pattern shows one quarter with concentrations twice those of the other three quarters, the range of possible bias in a mean is 88 percent to 136 percent if one-tenth of the total observations are in a single quarter.

Temporal balance is essential only if simple unweighted statistics are used. Unbiased estimates can be obtained from unbalanced samples by using appropriate sample stratification and by applying weighting factors. This method is particularly effective in determining unbiased estimates for a given sample size when the variance components are different within sampling strata.

The problems of unbalanced samples were considered in the establishment of the original validity criteria for TSP and other 24-hour data. The recommendations of Nehls and Akland for continuous data also considered these problems. Unfortunately, the current criteria for intermittent sampling do not ensure balance when sampling more frequently than once in 2 weeks; more stringent criteria or weighting factors need to be considered. In addition, the current NADB criteria for continuous sampling may not ensure balance while requiring more observations than necessary. Fewer observations could provide better estimates if the sampling recommendations were adopted.[2]

5.1.5 Estimates of Air Quality

The variation in a statistical estimate of air quality depends on the temporal variation of the data, the size of the data sample, the temporal distribution of the sample within the year, and how the data are combined in developing the estimate.

Data completeness has varying impacts on different sample statistics. In general, the more data available, the better the estimate. For a given level of uncertainty, however, fewer observations will be needed to estimate a mean than, say, a maximum value. For data with cyclical components, temporally balanced samples with less data will yield better estimates than unbalanced samples with more data. With the use of weighted averages, more data will usually yield better estimates, even if they are not balanced.

Intermittent Data - Nehls and Akland investigated the accuracy of annual means estimated from systematic samples. Table 7 shows how accuracy degrades with decreasing sampling frequency based on particulate data collected almost continuously in Philadelphia over 9 years. As expected, the percent error increases with decreasing sampling frequency. For a given sampling frequency, the error for a quarterly mean is approximately twice that for the annual mean.

TABLE 7. ACCURACY OF PARTICULATE SAMPLING FREQUENCIES AND AVERAGING INTERVALS

Sampling frequency (k) days	Percent error compared to daily sampling		
	Yearly	Quarterly	Monthly
2	1.4	2.3	3.9
3	1.7	3.2	6.5
4	2.3	4.6	8.7
5	2.7	6.0	10.3
6	3.2	6.5	12.0
7	8.7	10.7	15.1
8	3.8	8.1	14.9
9	5.0	8.9	16.0

Empirical evidence of the effect of sample frequency on summary statistics is presented in Table 8, based on TSP data from 14 sites monitoring continuously for 3 years, 1976-78. Estimates of annual averages and annual maximums when sampling every other day, every 3rd day, every 6th day, and every 12th day are compared with the averages and maximums obtained from daily sampling. Errors of the estimates for the annual maximum are typically 4 to 6 times larger than the errors in the mean. Even sampling every other day, the sample maximum underestimates the annual maximum by 9 percent.

TABLE 8. DEVIATION OF OBSERVED ANNUAL MEAN AND MAXIMUM FROM TRUE VALUES AMONG 42 TSP SITES, 1976-78

Sampling frequency	Average percentage error, (%)	
	Mean	Maximum
1 in 12	7.1	30.1
1 in 6	4.1	22.0
1 in 3	2.1	13.6
1 in 2	1.9	8.9

Continuous Data - The variability of an annual mean derived from continuous data can be examined by assuming that the data consist of 1 or 2 months of complete data in each calendar quarter. This assumption corresponds to the extreme situations in which the quarterly completeness is 33 percent or 67 percent.

Variance of the annual mean $Var(\bar{x})$ based on complete months of data is expressed as

$$Var(\bar{x}) = \frac{\sigma^2}{4n} (1-f) = \frac{\sigma^2}{4} \left(\frac{1}{n} - \frac{1}{3}\right), \quad (1)$$

where f is the fraction of months sampled in each quarter (f = n/3 and 1-f is the finite population correction), n is the number of months in a quarter, and σ^2 is the variance among true monthly means within a quarter. For simplicity, the number of days in a month and the number of months sampled per quarter are assumed to be equal. To generalize this approach, this model could be constructed to consider subsamples of any number of discrete blocks of consecutive days within each calendar quarter.

The variance of an annual mean based on monthly data defined by Equation 1 depends on the variability among monthly means within a quarter and on the number of months sampled in a quarter. Thus, for 1 month of data per quarter, the standard deviation (SD) of the mean is 0.41 times the SD of monthly means. For 2 months of data per quarter, the SD of the mean is 0.2 times the SD of monthly means.

The model based on monthly data was evaluated by comparing the empirical distribution of annual means (based on 1 or 2 months of actual data) with the theoretical distribution defined by Equation 1. Complete SO_2 measurements observed at the NYC laboratory site during 1973 were used to generate all 81 possible subsamples and corresponding annual means. Using the average quarterly estimate of variance among months equal to $(21.4)^2$, the estimates of variance for the sample means using 1 month or 2 months of data per quarter are $(8.8)^2$ and $(4.4)^2$ respectively. Normal distributions corresponding to these variance estimates were compared to the histogram based on all 81 possible annual averages; there was good agreement, as indicated in Figure 1. This example demonstrates that data with one-third of the observations in each calendar quarter can produce reasonable sample estimates.

The analysis above assumes that the variance is the same in each sampling strata (quarter). If variances are different, the sampling period with the highest variability should have the best sampling representation.

5.1.6 Summary and Recommendations

Data completeness criteria for producing summary statistics are desirable to ensure representative estimates of air quality. Ideally, different criteria

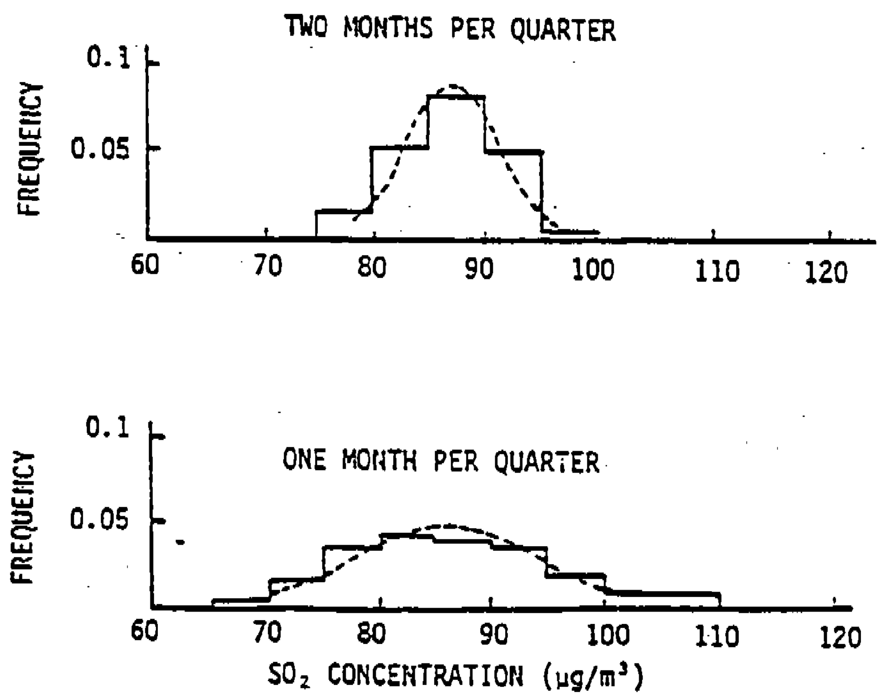


Figure 1. Empirical histograms of all possible annual averages based on one or two months of data per quarter compared with theoretical frequency distributions defined by statistical model for 1973 SO_2 at the New York City laboratory site.

should be used to screen data for different uses; however from a practical point of view, a single criterion specifying minimum data completeness for one statistic (e.g., a mean) will probably be used for all applications. The current NADB criteria for summarizing data need to be revised. The recommendations of Nehls and Akland appear reasonable for intermittent data; the existing criteria are reasonable as long as weighted statistics are used to correct for the potential sampling imbalance. Methods for specifying the required strata and numbers of observations should be resolved in future work.

5.2 HISTORICAL COMPLETENESS

5.2.1 Background and Purpose

Important to trend analysis is data base preparation. If one is analyzing the trend at only one monitoring site, all data from that monitor can be considered. When analyzing a group of monitors, a specific time frame represented by the group should be used, and the historical data completeness of each candidate monitoring site should be examined to determine suitability for trends analysis within the time frame.

Since air quality trend analysis focuses on year-to-year variation (as opposed to within-year variation), emphasis is usually on comparison of annual statistical indicators derived from complete or "valid" data sets. Thus, the number and temporal distributions of the annual statistical indicators are important; this characteristic is termed "historical completeness."

The data series should be as complete as possible, but missing values are permissible. Air quality monitoring does not always produce complete data records. Instrument failure and data processing problems are two of the many reasons for missing data. In many cases, emphasis is on compliance assessment rather than long-term monitoring. Historical completeness criteria are used during the data screening process to identify the largest possible data base from which a representative sample can be drawn for trend analysis.

5.2.2 Characteristics of Missing Data in Trend Analysis

The number of missing values permitted depends on the objective of the trend analysis, the trend technique, and the variance components of the data. In this context, variance components include error due to incomplete sampling, instrument error, meteorological fluctuations, and departures of the observed trend from an assumed underlying pattern. The sensitivity of a trend technique improves with more data; thus the analyst desires the maximum amount of data at the largest number of sites. The sensitivity also depends on the variability of the trend. If, for example, a monotonic trend were to exist at a group of sites, a minimum of two observations selected at random from each site would be sufficient to categorize the trend. Similarly, if the analysis objective were to detect a long-term shift without the need to categorize the year-to-year pattern, a few widely spaced data points would be adequate. If, however, the temporal patterns were more complex or were masked by high data variability, more data would be needed to separate the trend from other variance components, and more complete temporal records would be needed to obtain an accurate year-by-year trend.

5.2.3 Historical Completeness Criteria

At a minimum, each site in the trend analysis must have two data points representing the time series. To ensure that these data points are widely spaced, the time period may be divided into two segments. For example, if the 6-year period 1972-77 were divided into 1972-74 and 1975-77, a site could be selected if it produced one valid year of data in the first and in the second time segments; this procedure ensures that data from the start and end of the time period are represented.

A data selection approach often used by EPA is based on the historical completeness of quarterly data. A criterion commonly used in the analysis of national trends is four consecutive calendar quarters with valid data in each of 2 time segments. Because the last one or two quarters of data can be missing from the most recent years due to late reporting, this approach can yield more current estimates than the approach based on complete annual data. In addition, using quarterly data to derive annual summary statistics minimizes the bias caused by within-year sampling imbalance (Section 5.1).

Using historical completeness criteria can increase the number of candidate monitors many fold, and thus help obtain a more geographically representative sample. Of the 4000 TSP monitors reporting data to the NADB between 1972 and 1977, 2661 met the aforementioned criteria based on valid annual data, and 2737 met the criteria based on valid quarterly data. The trend period was divided into two time segments--1972-74 and 1975-77. Using three 2-year time segments and valid annual estimates, 1737 would qualify, but only 417 had valid annual data in each year between 1972 and 1977.

If a representative sample of stations with valid data for all years is available, additional sites with missing data may not be needed or may only be needed to improve the detectability of trends or to confirm the existence of trends by using a larger population.

5.2.4 Historical Completeness in Data Presentation

Historical completeness of data is particularly important for graphical presentations of trend results. Trend lines based on a group of monitoring stations are commonly used in air quality analysis. Such an aggregate trend line must be based on the same number of sites for each data point (e.g., year) to minimize bias caused by a changing data base; this requirement is satisfied by using every-year stations or by estimating all missing values. Estimating missing values is not necessary for trend assessment, but it is for data presentation.

In general, EPA characterizes the air quality trend in a defined geographic region using data for n years from m monitors. These data can be arranged in the matrix:

	Year 1	Year 2	. . .	Year n
Monitor 1	x_{11}	x_{12}	. . .	x_{1n}
Monitor 2	x_{21}	x_{22}	. . .	x_{2n}
. . .	. . .	. . .	x_{ij}	. . .
Monitor m	x_{m1}	x_{m2}	. . .	x_{mn}

where value x_{ij} occurs at monitor i in year j. The presentation of trends requires an average regional value of x for each year. If the matrix is complete, a reasonable estimate of the average value for year j is the arithmetic mean of each column,

$$\hat{\mu}_j = \frac{1}{m} \sum_{i=1}^{m} x_{ij}. \qquad (2)$$

However, this estimate may be biased if the matrix has missing values.

Two methods have proved useful in estimating yearly means when the data matrix is incomplete. In the one method, missing values are estimated by linear interpolation before the column means are calculated. This method produces reasonable estimates when the underlying trend is linear. The other method assumes each cell in the matrix can be characterized by a general linear model:

$$x_{ij} = \mu + M_i + Y_j + e_{ij} \tag{3}$$

where

μ is a constant--the grand mean of all the responses which would be obtained if there were no errors;

M_i is a term peculiar to the ith monitor, and is independent of year;

Y_j is a term peculiar to the jth year, and is independent of monitor; and

e_{ij} denotes the experimental error associated with x_{ij}, and it is assumed to be a normal random variable with standard deviation σ_e.

Computer programs can be used to generate least squares estimates of these parameters regardless of the number of missing values in the data matrix. Each yearly mean is estimated as

$$\hat{\mu}_j = \hat{\mu} + \hat{Y}_j. \tag{4}$$

Statistical methods are available for determining confidence intervals for the yearly means and for the differences between yearly means.

One advantage of the generalized linear model is that it explicitly relates the number of missing yearly values to the uncertainty of the estimates of the yearly means. Working backwards through the model, the analyst can specify a desired confidence interval, calculate the number of permissible missing values, and adjust the data base accordingly.

5.2.5 Effect of Historical Completeness on Trends Analysis

The impact of missing data on trends analysis is empirically examined by sampling from a group of 30 SO_2 sites with 75 percent complete data each year of 1974-78. Theoretical 90 percent and 50 percent probability intervals based on sampling one-third of the annual means in each of the first 3 years and one-half in each of the last 2 years are shown in Figure 2. This method of sampling is analogous to selecting sites independently for each year, and the intervals show the distribution of means in each year.

The smallest variability is in the last 2 years because of the larger number of observations used to calculate those means. The distribution of means

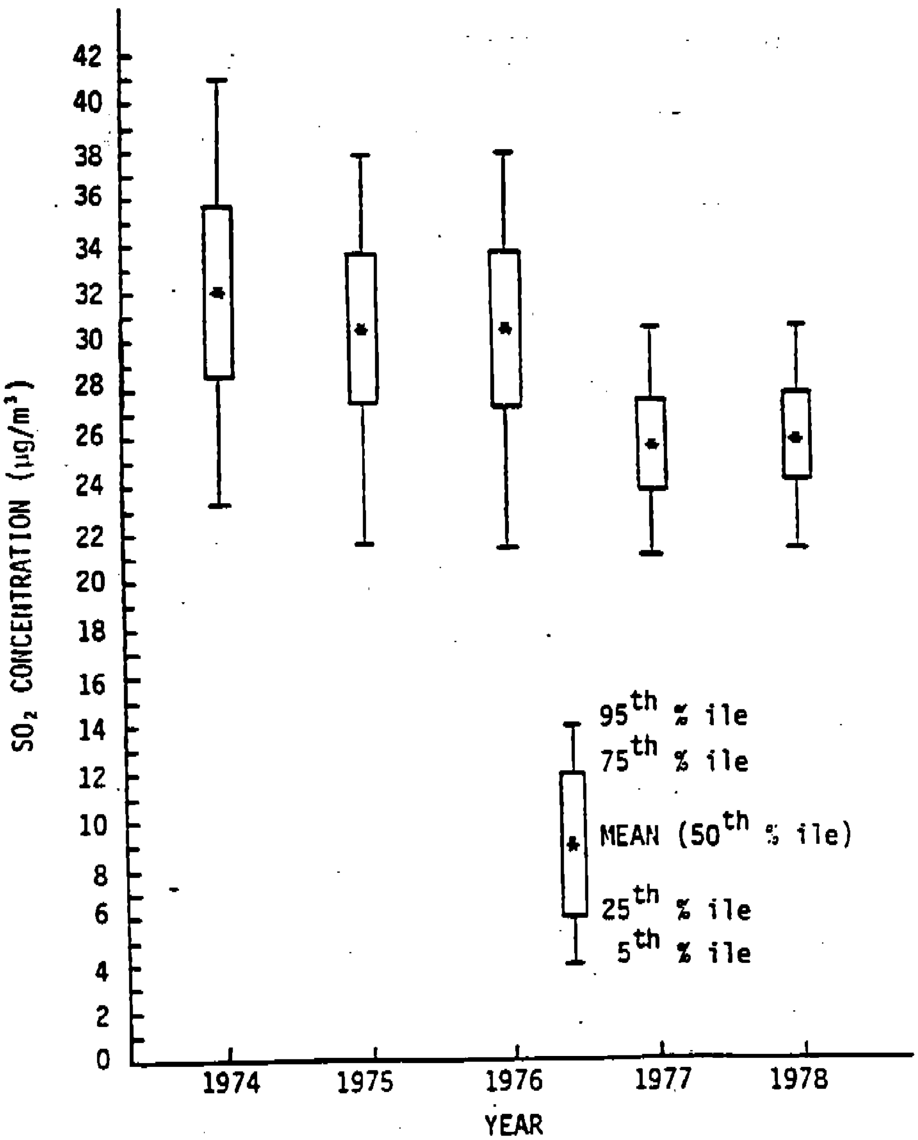


Figure 2. Theoretical probability distribution of annual mean SO_2 with 10 sites in each year 1974-1976 and 15 sites in 1977 and 1978.

is sufficiently variable for a subsample of means to show an incorrect upward trend. The extremes occur when a site effect exists and when different sites are used at the start and end of the time period. The highest and lowest annual means from the aforementioned subsamples are shown in Table 9.

TABLE 9. ANNUAL MEANS

	1974	1975	1976	1977	1978
Lowest	14.5	21.4	11.3	12.9	14.5
Highest	56.9	56.8	58.7	35.5	37.3

The situation is less ominous if an additional constraint--that the same sites be used in the first 3 years and in the last 2 years--is imposed as a requirement for historical continuity.

Sampling from a finite population dictates that the distribution of means for each year of the first 3 years will be independent of the distribution of means for 1 of the last 2 years. If there is a site effect, other combinations will be dependent. The most extreme results will occur if the distribution of means in the first and last years are independent; for example, there is a 25 percent chance that the subsample-derived mean of the first year will be less than the true value and that the subsample-derived mean for the last year will be greater than the true value. The extreme value for the first year could be 14.5 and that for the last year could be 37.3, but values for the intermediate years would counterbalance the extremes at the end years because of the site effect. Example values for the three intermediate years are 33.7, 42.2, and 14.9, respectively. Thus, historical continuity minimizes the chances of an incorrect result, but this rare combination of missing values may still prevent the detection of the true downward trend.

5.2.6 Conclusions and Recommendations

Historical completeness criteria are useful screening tools for selecting sites for trend analysis. Historical data should be as complete as the environmental data base will permit in order to establish a representative sample which is large enough to detect trends. Trend analysis should keep track of the extent of the historical completeness, and should report separate findings accordingly. Techniques such as linear interpolation and the generalized linear model should be applied to trend data with missing values to minimize bias due to site effects for the purpose of displaying graphical results.

5.3 REFERENCES

1. Aeros Manual Series, Volume III. Summary and Retrieval. EPA-450/2-76-009a, July 1977.

2. G. J. Nehls and G. G. Akland. Procedure for Handling Aerometric Data. J. of Air Pollution Control Association. 23:180 (1973).

3. Air Quality Data from the National Air Sampling Networks and Contributing State and Local Networks, 1964-1965. United States Department of HEW, 1966.

6. STATISTICAL INDICATORS AND TREND TECHNIQUES

This section attempts to answer two commonly asked questions: What is the air quality? How is it changing? Throughout the section, it is assumed that the principles and procedures of the previous sections have been followed; that is, the data sets have been screened for outliers and the sites have been examined to ensure representativeness. Separate subsections are devoted to each topic--statistical indicators and trend techniques, but a certain amount of interplay is involved between the two. For example, when a particular statistical indicator is used to summarize the data, a natural followup concern is how this indicator has changed over time.

6.1 STATISTICAL INDICATORS

The term "statistical indicator" is used in this section in a fairly general sense to include any statistic which summarizes air quality data for a particular time period. Technically, a statistic is a "summary value calculated from a sample of observations."[1] For some air quality applications, we may have not merely a sample but also the entire population; therefore, the computed value is not technically a statistic. For the purposes of this section, statistical indicator is used in either case.

This subsection discusses the background and purpose of statistical indicators for air quality, the types of indicators frequently used, what properties are desirable, and the relative merits of various indicators as well as future needs.

6.1.1 Background and Purpose

A continuous air quality monitor can produce a pollutant measurement for each hour of the day every day of the year. This means as many as 8760 concentration values for a single pollutant at a single site. The volume of data is further increased by the number of pollutants measured at a site and by the number of sites within an area. Such a quantity of data requires some type of data reduction to conveniently summarize the data. A wide variety of summary statistics could be used. The proper choice depends on the intended purpose of the information and on an awareness of the characteristics of the data.

While many specific purposes could be listed for air quality statistical indicators, for most practical applications there are basically two: (1) indicating status with respect to standards and (2) evaluating trends. An air quality standard is an absolute frame of reference for an indicator. If an air quality statistic is not compared to a standard, it is usually compared to a value from some other time period under the general heading of trend analysis. A third purpose is using the statistic as a basis of comparison with data from other sites or cities; however, most remarks that apply to choosing an indicator for trend analysis also apply to these comparisons.

An important point in any discussion of statistical indicators is that many air quality standards are structured in terms of concentration limits not to be exceeded more than once a year. This places a premium on information in the upper tail of the distribution. Accordingly, the average and median values commonly used as summary statistics in many fields are often of little interest in air quality data analyses. In fact, only the extreme values may be of interest for comparisons with standards.

6.1.2 Types of Statistical Indicators

A wide variety of statistical indicators are being used in air quality data analyses. Some reflect standard statistical choices such as means or percentiles. The highest or second highest value or the number of times the level of the standard is exceeded is often used due to the importance of the higher concentration values in air quality management. For example, the 1-hour NAAQS for CO specifies a level of 9 ppm not to be exceeded more than once a year. Thus, the second highest hourly value or the number of hourly values greater than 9 ppm would equivalently indicate whether or not the site meets the standard.

Other statistical indicators represent a compromise between focusing on the peak concentrations and introducing more stability into the indicator. These indicators may use either upper percentiles or knowledge of typical pollutant patterns to construct a useful index for trends. Examples include the average of daily maximum O_3 measurements for the O_3 season or the number of times a level (other than the standard) is exceeded.

Although the previously discussed indicators are typically used to summarize data for one pollutant at one site, they can be adapted to accommodate data for one pollutant at several sites within an area. A further generalization is to incorporate data for several pollutants into a single index. To do this requires some method of normalizing the individual pollutant measurements so that there is some basis for scaling their relative contributions to the index value. The Pollutant Standards Index (PSI) is a method for doing this.[2,3]

An ideal indicator would incorporate both air quality data and population data to provide a measure of population exposure to various air pollution levels. Efforts have been made in developing such measures, but the degree of spatial and temporal resolution required for both data sets makes these beyond the current state of the art. However, simplifying assumptions may be introduced to obtain rough approximations for these types of measures.[4]

6.1.3 Desirable Properties

Certain properties are desirable for a statistical air quality indicator. Some properties are desirable on a purely intuitive basis; others involve technical or practical considerations.

Clarity - Because the purpose of an indicator is to convey information, there is an advantage in using an indicator that is easily understood. Although understanding is partially a function of the intended audience, simple data presentations should use more easily understood indicators than those appropriate for a detailed examination of alternative control strategies if the more detailed analysis requires a more complicated indicator. Clarity does not necessarily mean that the indicator is simple to compute. For example, the computations for evaluating the PSI require linear segmented functions, but the final result is relatively easy to comprehend.

Independence of Sample Size - Another desirable property, on an intuitive basis, is that the indicator be independent of sample size. For example, sampling TSP every 6th day commonly results in approximately 60 measurements a year. Sampling every day of the year may result in 365 (or 366) measurements. Use of the maximum value, the second maximum, or the number of times the standard level is exceeded creates a problem with different sample sizes. A site that samples once every 6th day has only a 1-in-6 chance of measuring the annual maximum; a site that samples every day obviously measures the maximum. Any indicator that does not account for varying sample sizes can be misleading.

Robustness - A property that should be mentioned is robustness--that is, the indicator is not unduly influenced by a few outliers. How desirable this property ~~is~~ for air quality data analysis depends on the type of study. In air pollution work, the higher concentrations are often the most important; therefore, the median concentration may be robust but irrelevant. If the purpose of the study is to examine trends resulting from an overall emission reduction, the robustness of the indicator may be desirable.

Precision and Accuracy - A statistical indicator can have all of the above properties and yet fail to adequately summarize the data. Therefore, precision and accuracy must be included as desirable properties. Accuracy implies that the indicator is unbiased; precision measures the variability of the estimate. A recent paper by Johnson[5] discusses these properties for certain air quality indicators.

Feasibility - A practical evaluation of potential indicators must consider the feasibility of implementing a particular choice. An indicator may be highly desirable but not feasible to implement due to the present state of the art or to lack of information in the air quality data bases. This does not mean that a promising indicator should be ignored because it is difficult to implement; such a situation can highlight where additional work is needed.

Relevance - After all of the above properties have been considered, the final test of an air quality indicator is its relevance to the purpose of the study. Does the indicator sufficiently characterize the data so that the results may be easily translated into a clear statement of the information in the air quality data?

6.1.4 Discussion of Candidate Indicators

Consideration of all possibilities is not practical when discussing candidate indicators. It is more convenient to delineate certain classes of indicators: peak values (highest and second highest measurements), mean values (arithmetic and geometric), percentiles, exceedance statistics, design value statistics, multisite indicators, multipollutant indicators, and seasonal indicators.

Peak Value Statistics - This class includes indicators such as the maximum and the second highest concentrations which have been used because they are consistent with the "once per year" type of air quality standard. These are not independent of sample size. For less than everyday sampling, they have a negative bias because they underestimate the true value. Therefore ~~it is~~ difficult to recommend this class of statistics unless an adjustment is made for sample size. A possible alternative approach is the use of expected peak values estimated by fitting distributions to the data;[6] this approach yields statistics independent of sample size.

Mean Values - TSP, SO_2, and NO_2 have NAAQS's involving annual means (arithmetic for SO_2 and NO_2; geometric for TSP), so the use of an annual mean for these pollutants is fairly natural. For CO and O_3, the annual mean is not particularly useful for assessing status. Because the primary CO control strategy involves motor vehicle emission reductions, the annual mean may be adequate for trend analyses; nevertheless, changes in higher concentrations, should be examined. The typical diurnal and seasonal patterns for O_3 suggest that the annual mean is of little value.

Percentiles - Use of percentiles is one means of adjusting for differences in sample sizes and affording a degree of protection against a few extreme values. If higher concentrations are of interest, little will be gained by using lower or midrange percentiles. Upper percentiles are more useful; the 90th, 95th, and 99th percentiles provide an adequate range for trends analyses.

Exceedance Statistics - This class of statistics includes both the frequency and the relative frequency that an air quality level (NAAQS or other) was exceeded. This type of statistic is intuitively appealing, and is relatively easy to understand. If the level of the standard is used, the results relate directly to status assessment. Statistics involving the recorded number of exceedances (rather than the percentage of exceedances) depend on the sample size; with less than complete sampling, they underestimate the true value. As with peak value statistics, it would be advisable to use an adjustment to account for incomplete sampling. This type of correction was incorporated into EPA's recently revised O_3 standard.[7,8]

A change from one to two exceedances does not mean that the air quality has become 100 percent worse. Caution is needed in interpreting this type of statistic because it exaggerates percentage changes. Johnson[5] considered the relative precision of exceedances and the 90th percentile for O_3 data, and concluded that the 90th percentile was more desirable for trend analyses.

Design Value Statistics - For a site that fails to meet a standard, the design value is the concentration that must be reduced to the level of the standard for site compliance. Design values computed for sites that meet the standard would be less than or equal to the standard. The design value is a convenient summary statistic, but it is not easy to compute because it may involve factors other than the actual air quality concentrations.[8] Although informative, these statistics are difficult to recommend for general use because of the complexities of their estimations.

Multisite Indicators - An indicator that combines data from several sites (e.g., PSI) can be useful. A recent paper by Cox and Clark[9] suggests areawide indicators for examining trends for regional-scale pollutants such as O_3. A potential problem with multisite indicators is that data from cleaner sites can mask higher concentrations at other sites. The seriousness of this problem depends on the purpose of the study.

Multipollutant Indicators - An indicator that combines data for several pollutants should scale the measurements so that the contribution of each pollutant is appropriate. The PSI was developed for this purpose, and is recommended as the standard indicator for these types of applications.[2,3] It provides a convenient air quality indicator for daily reporting. A multipollutant indicator is not recommended for characterizing trends because of the difficulty of interpreting the results; for example, improvement in one pollutant may mask degradation in another. The logical step in interpreting the results is to examine trends for each pollutant.

6.1.5 Conclusions

When the purpose of the analysis is to compare standards, an indicator that relates directly to the standard is needed. For pollutants with annual mean standards, the annual mean also suffices for trend analyses. For pollutants with only peak value standards, peak value and exceedance statistics are acceptable for trends if adjusted for sample size; upper percentiles (90th, 95th, and 99th) should also be considered. For pollutants (e.g., O_3) with clear seasonal peaks, statistics based on data for only the peak season are acceptable. In all cases, the data should be summarized for an averaging time that corresponds to the averaging time of the standard.

Multisite indicators may warrant further study, particularly for areawide pollutants. In national trend assessments, data are available for hundreds of sites; but because many subsets of these sites are correlated, interpretation of the results is complicated. Multisite indicators could be used to aggregate the data for appropriate subregions as an intermediate step in evaluating trends.

Another area that warrants further attention is the development of population exposure indicators. Such indicators could provide not only a useful technical tool but also an effective means of conveying information to the general public.

6.2 TREND TECHNIQUES

A question of interest is whether air quality has changed over time. The search for the answer results in a variety of analyses categorized as trend assessment. Air quality trends analyses can vary in complexity from a simple numerical comparison of statistics from two time periods to a detailed time-series analysis incorporating the effects of meteorology and emission control programs.

This subsection is a general discussion of trend techniques for air quality data. More detailed information on certain techniques is contained in EPA's *Guidelines for the Evaluation of Air Quality Trends.*[10] Another useful report, *Methods for Classifying Changes in Environmental Conditions,* has recently been prepared in conjunction with the development of EPA's environmental profiles effort.[11]

6.2.1 Background and Purpose

Because of the general public's interest in air pollution, questions concerning air quality are fairly common. Air pollution summaries in units such as micrograms per cubic meter need a useful frame of reference to make the result more meaningful. The NAAQS's are one basis of comparison; another is how air quality values have changed over time, which is a convenient practical way to provide a perspective on what the numbers mean.

Both the general public and the technical community have a vested interest in air pollution control programs, and trends analyses are frequently used to evaluate the effectiveness of these programs. These analyses can be quite sophisticated. For example, ambient CO trends during the mid-1970's in New Jersey were the result of both the Federal Motor Vehicle Control Program (FMVCP) and an Inspection/Maintenance (I/M) Program instituted by the State; this result was further complicated by the residual impact of the gasoline shortage in 1974-75.

These factors, in addition to the possible influence of varying meteorology, rapidly escalated the level of detail needed in an analysis--not only with respect to the trend techniques used but also with respect to the types of data needed.

Another consideration in air quality trends analyses is the scope of the study. The study may involve data from a single site, or it may involve a national data sample from hundreds or thousands of sites. A technique that may be reasonable for an individual site may not be feasible for a more extensive data set.

6.2.2 Types of Trend Techniques

A wide variety of trend techniques have been used for air quality trends analyses. This section briefly discusses nonstatistical and statistical approaches.

Nonstatistical - The usefulness of graphs (discussed in more detail in Section 8) in any data analysis should not be underrated. At its simplest level, the purpose of a trend analysis is to determine what patterns are present in the data. Box-plots,[12] time-series plots, histograms, and so forth--all serve as convenient guides to the data analyst. In many cases, a careful choice of the proper graphical technique suffices to indicate trends (e.g. Whittaker Henderson,[13] Box-plots,[14,15] and two-way tables[16,17,18]).

A second nonstatistical approach is a simple numerical comparison between summary statistics for two time periods. This is sometimes modified so that trends are categorized as "no change" if the change fails to exceed some specified limit. This seemingly arbitrary cutoff may involve an underlying statistical rationale and an assumed variance component.

Statistical - The statistical technique may merely be a comparison of two annual means by using a standard t-test. When a sequence of data values is available, regression analysis may be used. This may be either parametric[19] or nonparametric.[14,15] Trends analysis using regression basically assumes a linear trend over time. In many practical situations, this assumption may not be reasonable, so alternative procedures (e.g., analysis of variance (ANOVA)) may provide a more flexible framework for indicating change. Because of the time dependencies in sequences of air pollution data, time-series models have been used for air quality trends analyses. Intervention analysis is one of the techniques used to examine control strategies.[20]

The above techniques apply to a single series of data points. For analyses of data from several sites, the results at each site may be summarized as either up or down, and then analyzed as a contingency table; this technique assumes that the sites are independent.

6.2.3 Desirable Properties

From a practical viewpoint, a desirable trend technique would be one which is intuitively easy to understand, feasible to implement, based on reasonable assumptions, and capable of providing results that are easily interpreted.

Intuitively Easy to Understand - People are more comfortable if the rationale for the analytical technique is intuitively understandable. A technique may always be implemented on a "black box" basis: the user submits the required input data, and the answer appears as the output. For the result to be useful, however, the user must be confident that the technique will work. This does not necessarily mean that the technique itself must be simple; it means that the underlying concepts must be easily explained. For example, an autoregressive, integrated, moving average model can be presented with a barrage of notations involving forward and backward difference operators. Yet, the need for a model that incorporates seasonal and diurnal patterns and dependencies from one hour to the next is more easily understood.

Feasible to Implement - It is a truism to state that the best trend technique available will be ignored unless it is feasible to implement. For air quality analyses, several considerations are involved in evaluating feasibility. Because there are limited data histories in many practical situations, the technique must be applicable to relatively short time periods (e.g., 2 to 5 years). For an analysis of data from a few sites, one may compute many types of summary statistics; for national analyses, choices must be limited to those readily available from NADB, which essentially limits the selection to certain standard quarterly or annual statistics. (Monthly statistics are not stored, and would have to be computed from raw data.) The number of sites involved also places practical constraints on the amount of computer time and analyst time available per site.

Based on Reasonable Assumptions - Every statistical procedure has an underlying model requiring that certain assumptions be satisfied. These assumptions should be met by the air quality data for a test to be useful; therefore, the stipulations should be structured so that the air quality data can reasonably be expected to meet them.

Results Easily Interpreted - After the analysis is complete, the success of a technique depends on whether the results can be easily interpreted. This is one advantage of a statistical technique, as opposed to a nonstatistical technique, for assessing change. The probability assigned by the statistical procedure is an aid in interpreting the results; it also affects the type of statistical technique that should be used. For instance, if the annual rate of change is the variable of practical interest, a procedure such as regression should be used since this is the type of information it produces.

6.2.4 Discussion of Candidate Techniques

Important in any discussion of candidate trend techniques for air quality data is the interpretation of the final result. Because of the strong seasonality of certain pollutants (e.g., O_3), the question of interest often involves trends for a particular season. From a practical viewpoint, if the effectiveness of a control strategy is of concern, only the peak months of the year may be of interest. Because the NAAQS's are stated in terms of annual status assessments (with the exception of Pb, which is quarterly), analyses are often in terms of annual summary statistics. Use of an annual summary statistic reduces the number of data points available for the analysis; but if the trend in this summary statistic is the item of interest, the analysis should be structured around the statistic.

The remainder of this subsection briefly discusses broad classes of trend techniques, and indicates comparative strengths or weaknesses.

Graphical Procedures - Graphical presentations (discussed in more detail in Section 8) are included here to emphasize that they are a useful first step in any data analysis, not merely a final step for presentation purposes.

Numerical Comparisons - A simple comparison of summary statistics from two time periods is of minimal use in most cases because it provides no frame of reference for what is a normal change from one year to the next. Any underlying statistical rationale for categorizing changes as up, down, or as no change should be stated in the analysis. If the results of these comparisons are aggregated over many sites for a contingency table analysis, they are acceptable.

Statistical Comparisons - Statistical comparisons may be made using either parametric or nonparametric techniques. A typical parametric technique would be a t-test to compare 2 years of data, although the presence of seasonality may artificially inflate the error term and reduce the sensitivity of the test. Nonparametric tests may be used to avoid the assumptions that the errors are normally distributed. The Wilcoxon signed-rank test allows for seasonality, and is therefore a useful nonparametric technique. Aligned-rank tests have been suggested for assessing changes in environmental data;[11] however, experience with these tests for ambient air quality trends is limited at this time.

Regression - Both parametric and nonparametric regressions provide an estimate of the rate of change as well as a statistical significance test. In assessing air quality trends, the rate of change is often important because it relates directly to the effectiveness of a control strategy rather than simply stating whether an increase or decrease has occurred. Because of seasonal patterns in the data, regression is normally applied to an annual statistic or a summary statistic for the peak season; this is a disadvantage since only one value a year is used. Regression also tests for a linear trend, which is useful either when the user is interested in a linear trend or in a relative measure of net change over the time period; but if a finer resolution of the pattern is desired, regression may not be adequate.

Analysis of Variance - The general class of analysis of variance (ANOVA) models offers considerable flexibility for trends analysis. The user can introduce month, season, and year terms into the model if a single site is being examined. For an areawide analysis, a site-effect term can be added as well as interactions and even covariates. ANOVA is not restricted to linear trend assumptions, and multiple comparison tests are available to test for significant differences among effects of interest.

Time-Series Analysis - A basic assumption in many statistical techniques is that successive measurements are independent--that is, the value of a particular measurement does not depend on past measurements. In view of the diurnal and seasonal patterns often present in air quality data, one may use time-series techniques for these problems. Applications of time-series techniques for air quality analyses have used Box-Jenkins techniques,[20] Fourier analysis,[21] and polynomials to remove the seasonal components.[22] Intervention analysis techniques seem appropriate for examining control strategies and other types of intervention.[20] While these techniques have the advantage of producing more information, they are also more difficult to apply; consequently, it is difficult to recommend them for routine large-scale analyses.

6.2.5 Conclusions

At the present time, it does not seem feasible to recommend specific trend techniques. Since trends analyses are done for different purposes, the level of resolution required in the answer may vary from a simple up-down classification to an actual determination of the percentage improvement associated with a specific control strategy. However, it does seem advisable to encourage the use of statistical tests as an objective means of classifying trends.

Areas that warrant further examination include the intervention/time-series approach to see if the implementation of this type of analysis can be simplified so that its use can become more routine. Also, more experience is needed in applying the aligned-rank test to air quality data.

6.3 REFERENCES

1. Kendall, M. G., and W. R. Buckland. A Dictionary of Statistical Terms. Hafner Publishing Company, Inc., New York, N.Y., 1971.

2. Guideline for Public Reporting of Daily Air Quality Data - Pollutant Standards Index (PSI). No. 1-2-044. Office of Air Quality Planning and Standards, U. S. Environmental Protection Agency, Research Triangle Park, N.C., August 1976.

3. Federal Register 44(92):27598, May 10, 1979.

4. Frank, N. H., et al. Population Exposure: An Indicator of Air Quality Improvement. 70th Annual Meeting of the Air Pollution Control Association, Toronto, Canada, June 1977.

5. Johnson, Ted. Precision of Quantile and Exceedance Statistics. American Society of Quality Control Technical Conference Transactions, Atlanta, Ga., 1980.

6. Johnson, T., and M. Symons. Extreme Values of Weibull and Lognormal Distributions Fitted to Ambient Air Quality Data, Paper No. 80-71.4, Annual Meeting of the Air Pollution Control Association, Montreal, Canada, June 1980.

7. Federal Register 44(28):8282, February 8, 1979.

8. Curran, T. C., and W. M. Cox. Data Analysis Procedures for the Ozone NAAQS Statistical Format. Journal of the Air Pollution Control Association 29(5):532, May 1979.

9. Cox, W. M., and J. B. Clark. An Analysis of Ozone Concentration Patterns Among Eastern U.S. Urban Areas (draft).

10. Guidelines for the Evaluation of Air Quality Trends, OAQPS No. 1.2-015. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C., February 1974.

11. Methods for Classifying Changes in Environmental Conditions. VRI-EPA 7.4-FR3 0-1 (draft). Vector Research, Inc., Ann Arbor, Michigan, February 1980.

12. Tukey, J. W. Exploratory Data Analysis. Addison-Wesley Publishing Company, Reading, Mass., 1977.

13. Spirtas, R., and H. J. Levin. Characteristics of Particulate Patterns 1957-1966. U.S. Department of Health, Education, and Welfare, National Air Pollution Control Administration, Raleigh, N.C., March 1970.

14. National Air Quality and Emissions Trends Report, 1976. Publication No. EPA-450/1-77-002. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C., December 1977.

15. National Air Quality and Emissions Trends Report, 1977. EPA-450/2-78-052. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C., December 1978.

16. Phadke, M. S., et al. Los Angeles Aerometric Data on Oxides of Nitrogen 1957-72. Technical Report No. 395. Department of Statistics, University of Wisconsin, Madison, Wisc., 1974.

17. Tiao, G. C., et al. Los Angeles Aerometric Ozone Data. Tech. Report No. 346. Department of Statistics, University of Wisconsin, Madison, Wisc., 1973.

18. Tiao, G. C., et al. Los Angeles Aerometric Carbon Monoxide Data. Tech. Report No. 377. Department of Statistics, University of Wisconsin, Madison, Wisc., 1974.

19. The National Air Monitoring Program: Air Quality and Emissions Trends. Annual Report Volume 1. Pub. No. EPA-450/1-73-001a. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C., July 1973.

20. Box, G. E. P., and G. C. Tiao. Intervention Analysis with Applications to Economic and Environmental Problems. Journal of the American Statistical Association 70:708-79, 1975.

21. Phadke, M. S., M. R. Grupe, and G. C. Tiao. Statistical Evaluation of Trends in Ambient Concentrations of Nitric Oxide in Los Angeles. Environmental Science and Technology, Vol. 12, No. 4, April 1978.

22. Horowitz, J., and S. Barakat. Statistical Analysis of the Maximum Concentration of an Air Pollutant: Effects of Autocorrelation and Non-stationarity. Atmospheric Environment 13:811-818, 1979.

7. INFERENCES AND CONCLUSIONS

This section focuses on the problem of identifying cause-and-effect relationships after a statistically significant trend or difference between data sets has been determined. Previous sections have discussed the problems of data completeness, historical trends criteria, precision and accuracy of data, screening data for outliers, siting and representativeness, statistics for analyzing air quality data, and trend techniques. Assuming that all of these problems can be resolved in a particular analysis, the problem of interpretation remains.

7.1 BACKGROUND

At the present time, most air quality data available for analysis are submitted to NADB. The major enhancement to this system is the current designation of NAMS--a refined national monitoring network in areas with large populations and high pollutant concentrations. Each NAMS will meet uniform criteria for siting, quality assurance, equivalent analytical methodology, sampling intervals, and instrument selection to assure consistent data reporting among the States.[1] Precision and accuracy data will be available with the air quality data for the first time.

Though the enhancements achieved through the monitoring regulations will provide a much sounder national air quality data base for decisionmakers, the data are still seriously deficient in determining cause-and-effect relationships between air quality, emission changes, meteorology, and the impact of unanticipated events such as fuel shortages. Explaining the causes of air quality changes is difficult without information to supplement basic air quality data. For example, what is the impact of a local, regional, or national gasoline shortage on ambient air pollution levels? What is the probable impact of fuel switching or tampering with automotive emission control devices on air quality levels?

A case in point is the recent controversy over O_3 trends in Los Angeles. A significant increase was observed in O_3 levels between 1977 and 1978. An attempt was made by the South Coast Air Quality Management District[2] to determine how much of the increase was due to meteorology; they used a meteorological index developed by Zeldin;[3] stated that a significant increase in O_3 levels remained after they had adjusted the data for meteorology; and speculatively attributed it to the breakdown of catalytic control devices. What was needed was concomitant information on emission changes; independent estimates of fuel switching, tampering, and so forth; and more complete meteorological data.

The problem in making inferences and conclusions is largely the lack of supplemental information needed to interpret air quality data analysis and trends. This lack could be improved in two ways: (1) the data analyst could try to find concomitant information (e.g., weather data collected at airports by the National Oceanic and Atmospheric Administration), or (2) a long-term solution would be to apply the principles of experimental design and enhance the NAMS in two or more major urban areas with supplemental information on meteorology and emission changes. The first approach is discussed in the following section on case studies, and the second is discussed in the section on long-term solutions.

7.2 CASE STUDIES

Over the years, a number of attempts have been made to explain ambient air quality trends in terms of changes in emissions, meteorological conditions, or instrument measurement practices (e.g., changes in laboratory procedures or calibration). The approach can best be illustrated by examples, as in the following four case studies.

7.2.1 Comparison of SO_2 Trends and the %S Content Regulations in Distillate and Residual Fuel Oil in Bayonne, N.J.

The impact of regulations for controlling the sulfur content of fuels on ambient SO_2 levels was illustrated in Bayonne, N.J., by comparing ambient SO_2 levels with the effective dates of regulations limiting the sulfur content[4] (Figure 3). The EPA report[4] which presented the analysis stated that the improvement in SO_2 air quality at this site can be attributed primarily to regulations which became effective in New Jersey and New York during 1968-72. The trend at the Bayonne site is consistent with the national trend in SO_2 during this time period,[4] and is probably due to changes in the allowable sulfur content of fuel.

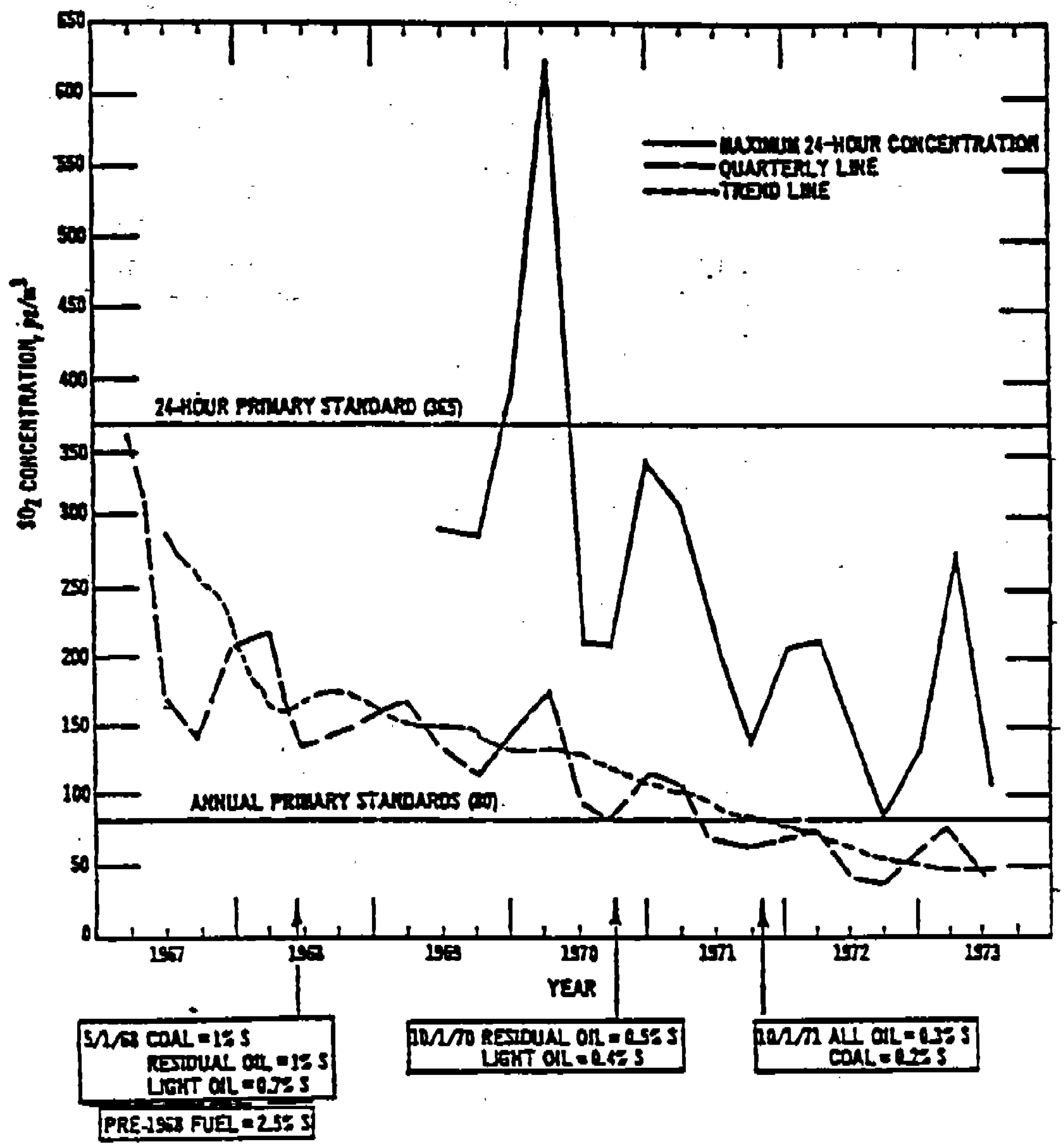


Figure 3. Comparison of SO_2 trends at Bayonne, N.J., with regulations governing % sulfur content in fuel.

The analysis could have been improved, and the conclusions strengthened if appropriate meteorological data had been available to be treated as a covariate. As it stands, however, the graphical presentation suggests a reasonable cause-and-effect relationship between the decreasing levels of SO_2 and the increasing restrictiveness of regulations limiting the sulfur content of fuel.

7.2.2 Impact of Gasoline Shortage on CO in Richmond, Va.

An analysis of CO data in Richmond, Va., illustrates the problem of interpreting ambient data. In 1974, an EPA analysis[5] was undertaken to evaluate the effect, if any, of the energy crisis on CO levels. The CO monitoring site in downtown Richmond was representative of the influence of commuter traffic patterns.

The period of time chosen for the analysis was the last 4 weeks (28 days) of January and the 4 weeks of February. This time period was expected to reflect the most severe period of the gasoline shortage and to minimize the potential anomalies in traffic patterns associated with the Thanksgiving, Christmas, and New Year's holidays. The windspeed data recorded at the R. E. Byrd International Airport served as an approximate indicator of the windspeed at the site. The data were presented as weekly averages to compensate for the differences in daily traffic patterns (Table 10).

TABLE 10. WEEKLY AVERAGE CO CONCENTRATIONS (ppm) AND WINDSPEEDS (mph) IN RICHMOND, VA, JANUARY 4-FEBRUARY 28, 1974

Week	Hourly average	Daily 8-h max.	6 to 9 a.m. and 4 to 7 p.m. average	Average daily windspeed
1/04-1/10	2.92	4.45	4.50	5.90
1/11-1/17	1.94	3.13	3.18	7.87
1/18-1/24	3.07	4.79	4.52	6.63
1/26-1/31	2.88	4.56	4.20	6.21
2/01-2/07	2.28	3.54	3.56	7.24
2/08-2/14	2.31	3.76	3.40	8.06
2/15-2/21	2.08	3.09	3.00	9.13
2/22-2/28	1.73	2.81	2.82	9.67

The CO data showed a downward trend during this 8-week period; however, the average windspeed increased during this period (Figure 4). The decline

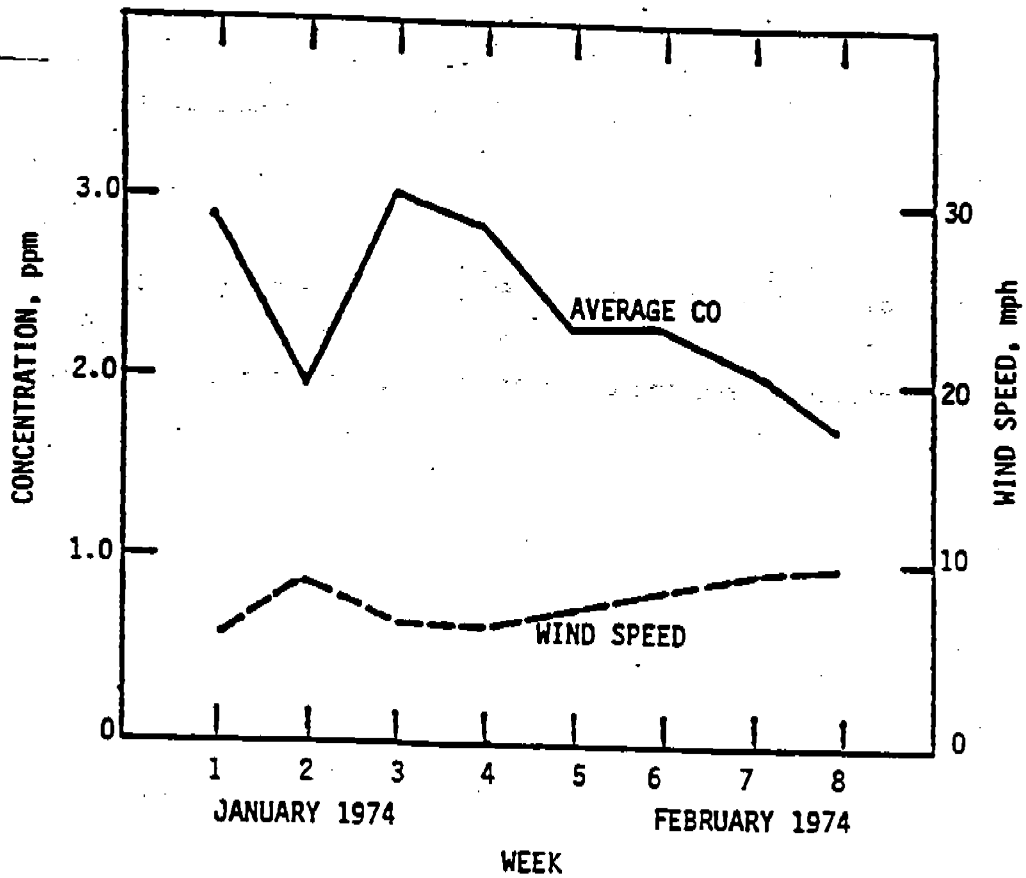


Figure 4. Weekly average CO and wind speed in Richmond, VA from January 4 to February 28, 1974.

in CO levels may, therefore, have been due to the increase in average windspeeds, which is indicative of greater dilution.

The interdependence was seen in the statistical analysis of the data. If the change in windspeed was ignored, four parameters--daily average, maximum 8-hour average, rush-hour average, and nonrush-hour average--showed statistically significant decreases, although the rush-hour decrease was less apparent in the averages. These findings were based on ANOVA. When windspeed was introduced into the analysis as a covariate, using analysis of covariance, none of the changes in the above parameters were significant. Therefore, although the data at this site indicated a decline in CO levels during this period, the associated increase in windspeeds made the cause of the decline difficult to assess.

Failure to detect significant trends after adjusting for windspeed is not entirely unexpected. The variability associated with CO measurements and the relatively brief duration of the gasoline shortage make it possible for the effect to go unnoticed, unless the monitoring site itself was in precisely the right location to detect changes due to alterations in traffic patterns.

This example illustrates one problem in trying to use data collected for one purpose for yet another purpose. The assumption that the windspeed at the airport reflects the windspeed at the downtown Richmond site may not be correct. Ideally, the windspeed should be measured at the CO monitor so the analyst could be more confident in interpreting the results.

7.2.3 Trends in CO in New Jersey

An analysis of CO data in New Jersey illustrates an attempt to relate statewide CO trends to gasoline consumption. Figure 5 was prepared by the N.J. Department of Environmental Protection to illustrate the progress made in reducing State CO levels from 1972 through 1976.[6] The dates for the initiation of the two phases of their inspection/maintenance (I/M) program are shown. New Jersey indicated that CO levels continued to improve despite an overall increase in gasoline consumption.

While progress in CO levels is clearly evident, the effectiveness of the New Jersey I/M program is confounded with the overall effectiveness of the CO emission reductions attributed to the Federal Motor Vehicle Control Program (FMVCP). Ideally, it would be desireable to separate the effects of the two control programs; this could be accomplished with a designed experiment. Concomitant meteorological data would be desirable to determine if changes in meteorology would aid in explaining the statewide CO trend.

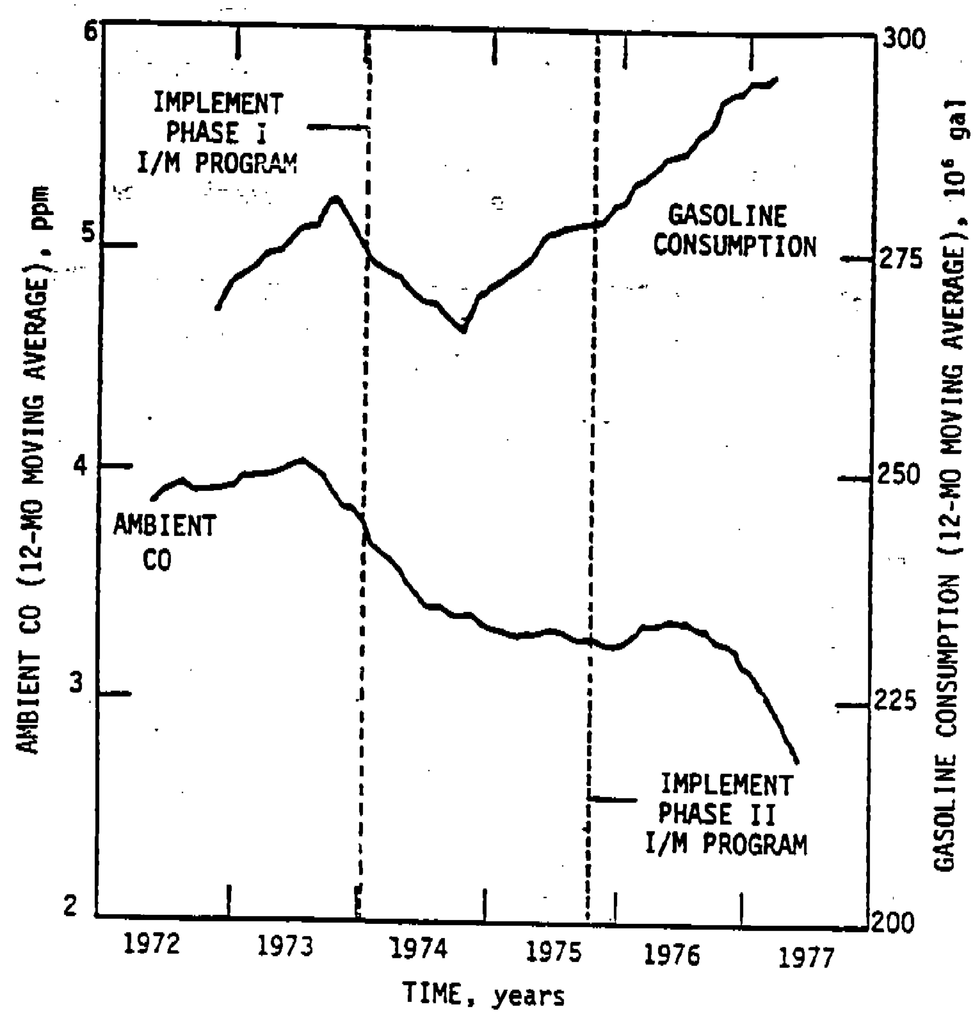


Figure 5. CO air quality from 18 monitoring sites and motor-vehicle gasoline consumption for N.J. from 1972 through 1976.

7.2.4 The Central Plains Drought

On February 24, 1977, the extremely dry soil conditions in the Central Plains and a strong frontal system resulted in dust being stirred up and transported east. Figure 6, a modified box-plot, shows peak TSP levels in Region VI (Central Plains) by quarter from 1972 through 1977; the dramatic increase in the first quarter of 1977 is obvious on this graph. Monitoring sites throughout Texas, Oklahoma, and Arkansas reported high TSP levels during this February duststorm. Several sites recorded daily values of more than 1000 $\mu g/m^3$; a single value of this magnitude would increase the annual geometric mean at a site by 10 percent.

In this example, the high TSP levels reported throughout Texas, Oklahoma, and Arkansas are substantiated by satellite pictures taken February 23-25, 1977 (Figure 7).

7.2.5 Conclusions and Recommendations

With the existing NAMS network, the best an analyst can do to facilitate the interpretation of air quality monitoring data is to seek other sources of information to help explain why an air quality trend has or has not taken place or why there is a significant difference between sets of data. The four case studies illustrate the importance of meteorology, emission changes resulting from control programs, and the impact of extraordinary events (e.g., the dustbowl of February 1977). A long-term solution is application of the principles of experimental design to the collection of air quality and appropriate concomitant meteorological data.[7] Such a solution is discussed in the following section.

7.3 LONG-TERM SOLUTIONS

One solution to the problem of trying to assess the effectiveness of EPA's control programs would be to select two or more urban areas and to collect concomitant information on a continuing basis to aid in answering both anticipated and unanticipated questions about effectiveness. The objective would be to determine cause-and-effect relationships between air quality and emissions after isolating the effects of meteorology; these relationships could be determined if data could be collected to estimate the impact of gasoline shortages, fuel switching, automobile control device tampering, contributions of various sources,

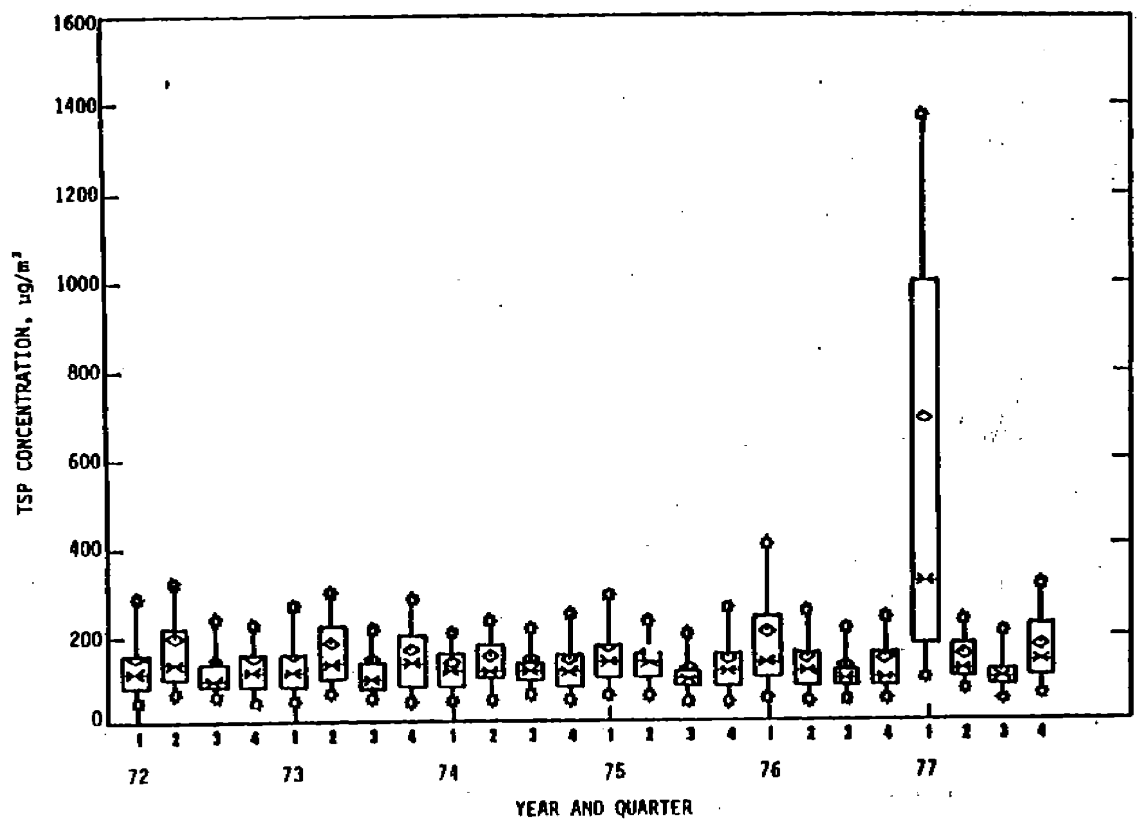


Figure 6. Quarterly TSP maximum values in Region VI from 1972 to 1977 illustrating the effect of the 1977 duststorm.

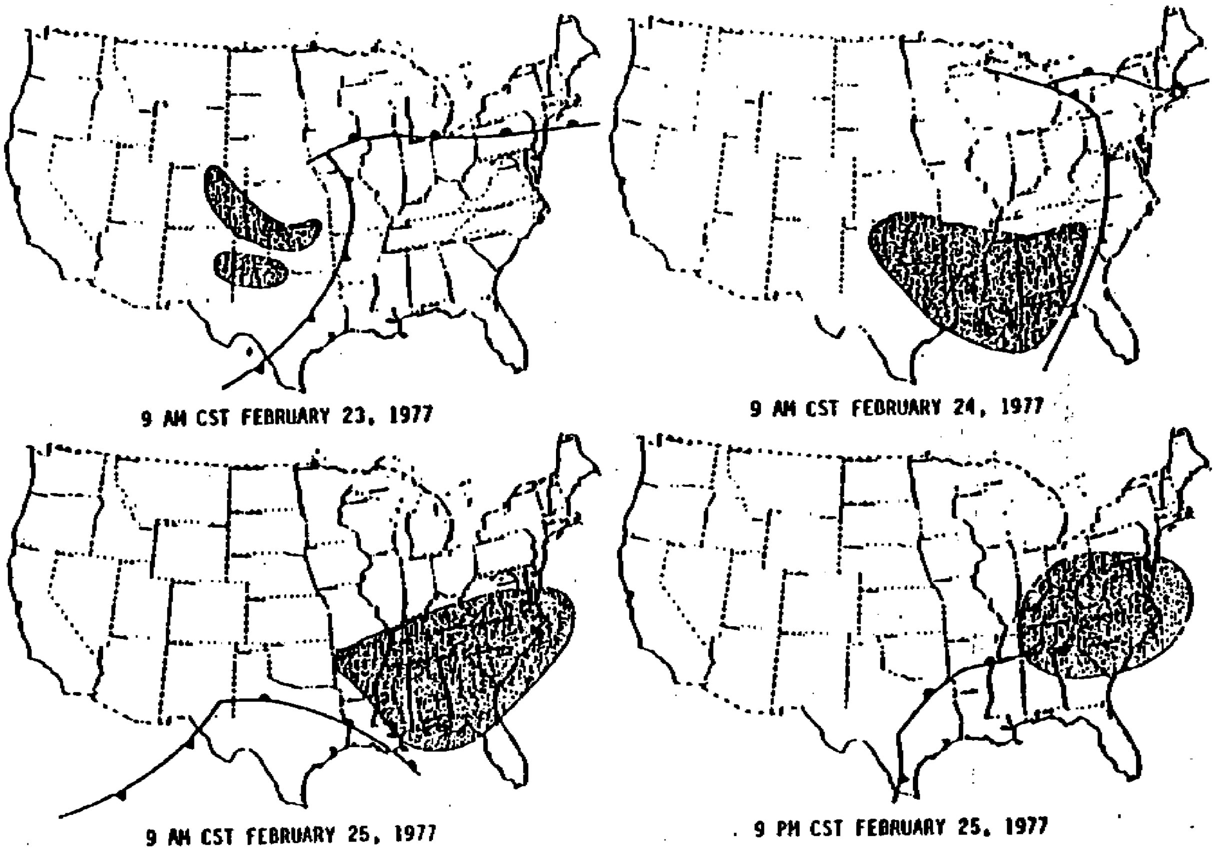


Figure 7. Satellite views of February 23-25, 1977, duststorm at succeeding time periods.

changes in instrument calibration, degree of human exposure, extent of fugitive dust, and so forth.

Air quality, monitoring, and emissions data would be collected according to an experimental design to test stated hypotheses. Each of the criteria pollutants--TSP, SO_2, CO, O_3, and NO_2--would be measured in the two or more urban areas along with IP and sulfates. Sources of emission changes to be considered in developing the experimental design include but are not limited to: (1) changing economic conditions; (2) increased or decreased refinery capacity; (3) fuel switching; and (4) growth patterns (e.g., the number of dry cleaning establishments).

At a minimum, experimental designs would use analysis of variance (ANOVA) and the analysis of covariance (COVA) to test hypotheses, and would use confidence intervals about means in question to display significant findings. Other appropriate statistical techniques would also be used to test hypotheses.

A contract has been let to explore this long-term solution in greater detail. The objective is to pose policy questions and then to determine what supplemental information would be needed to answer the questions in two or more urban areas. The associated costs are also to be determined. When the report on this contract is completed, its principal findings will be summarized and appended to this report.

7.4 REFERENCES

1. Federal Register. Vol. 44, May 10, 1979, pp 27588-27604.

2. Davidson, A. et. al., Air Quality Trends in the South Coast Air Basin. South Coast Air Quality Management District, El Monte, Calif. June 1979.

3. Zeldin, M. D., and D. M. Thomas, Ozone Trends in the Eastern Los Angeles Basin Corrected for Meteorological Variations. Presented at the International Conference on Environmental Sensing and Assessment, Las Vegas, Nevada, 1975.

4. Monitoring and Air Quality Trends Report, 1972. Pub. No. EPA-450/1-73-004. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C. December 1973.

5. Monitoring and Air Quality Trends Report, 1973. Pub. No. EPA-450/1-73-007. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C. October 1974.

6. New Jersey Department of Environmental Protection Annual Report: July 1, 1976-June 30, 1977. New Jersey Department of Environmental Protection, Trenton, N. J. (in preparation).

7. Use of Meteorological Data in Air Quality Trend Analysis. EPA-450/3-78-024. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C. May 1978.

8. DATA PRESENTATION

The purpose of this section is to review well-established data presentation techniques applicable to aerometric measurements and to provide guidance on selection of the forms most suitable to the scope of the problems to be analyzed. Selection is based on criteria such as audience applicability, spatial and temporal classification, and availability of computerized statistical and graphical resources. Therefore the discussion is focused on (1) the fundamental concepts to be displayed, (2) chart types, (3) input data forms, (4) statistical classifications, (5) audiences, (6) caveats/enhancements, and (7) available plotting resources. Finally, displays that meet these requirements are included.

8.1 CONCEPTS TO BE DISPLAYED

A simple but accurate presentation of statistical measures of large data bases or parametric relationships contributes greatly to the reader's understanding of the data. The most common displays of data on air quality and source emissions include (1) current status of the pollution problem, (2) statistical or descriptive trend, (3) impact of one or more parameters on another, (4) comparison of two or more groups or classes for a specific parameter, (5) relation of component parts to the whole, and (6) spatial and temporal patterns.

Statistical measures may be chosen not only from individual monitoring-site or emission point-source populations of averages, medians, percentiles, maximums or standards exceedances but also from aggregates (weighted or unweighted) of these sites on a city, county, region, state, national, or global scale.

Data presentation should be carefully planned. The charts or graphs, as well as the statistical displays, may be easily distorted (intentionally or unintentionally) by compression or expansion of scales, by nonuniform or broken gridding, by line and shading optical illusions, and by cluttered or complex information. The following subsection introduces the most frequently used displays.

8.2 CHART TYPES AND USES

8.2.1 Tabular Summaries

Summarizing data in tabular format is an important first step in data analysis. The validity and accuracy of the data should be certified before applying summary techniques. The tabular summary provides a permanent record of the individual data items which can be further analyzed in future research projects.

Aerometric examples of this form are NADB's monthly listings of hourly air quality values and yearly listings of intermittent (daily or composite) values. These data are also summarized by NADB in the yearly and quarterly frequency distribution listings. Similarly, point-source emissions data are submitted to NADB by the National Emissions Data System (NEDS).

Tabular summaries provide comparisons of several parameters categorized at one or more levels. Examples are listings of TSP, SO_2, and NO_2 by day of the year (Table 11), weekly TSP/maximum values at several monitoring sites (Table 12), and TSP monthly averages by site type (Table 13).

Finally, either certain qualitative measures may not be readily applicable to graphing or more information is provided in a smaller document. One example is a display of current attainment status and trends using a variety of symbols and colors, as shown in Figure 8.

Tabular listings or summaries should be limited in most cases to technical support documents. For regional or national publications, these summaries should be limited to comparisons of one or two parameters and about five classes or categories. Tedious or complex tables often contribute to the reader's confusion and misunderstanding; moreover, descriptions of statistical techniques used in producing these tables are often either left out or limited in the text discussion.

8.2.2 Point Charts

Point charts are used primarily to describe situations where scatter or clustering are important. Since these displays for large data bases would be almost impossible to draw manually, computerized techniques have been applied. Rapid display of data to detect outliers, parametric relationships, or data distribution is clearly a benefit to the statistician. Common uses are correlation or cluster analysis (Figures 9 and 10), plotting air quality concentrations by year (Figure 11), and comparison of air quality for several

TABLE 11. MULTIPLE PARAMETER LISTING, 1979, $\mu g/m^3$

Yr/mo/day	TSP	SO_2	NO_2
79/01/01	125	35	45
79/01/02		43	52
79/01/04	86	50	
.	.		.
.	.		.
.	.		.
79/12/31	145	83	96

TABLE 12. WEEKLY TSP MAXIMUMS AT CITY SITES, 1979, $\mu g/m^3$

Week beginning	Sites			City max	City avg
	4th & Market	6th & Vine	18th Jackson		
Jan 7	120	50	70	120	80
Jan 14	163	No sampling	99	163	131
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
Dec 29	103	52	63	103	73
1979 Max	236	99	136	236	187
1979 Avg	125	55	86	125	77

TABLE 13. REGION 5 MONTHLY AVERAGES BY SITE TYPE, 1979, $\mu g/m^3$

Month	Commercial			Industrial			Residential			Remote		
	Sites	Obs	Avg	Sites	Obs	Avg	Sites	Obs	Avg	Sites	Obs	Avg
Jan												
Feb												
.												
.												
.												
Dec												
Total												

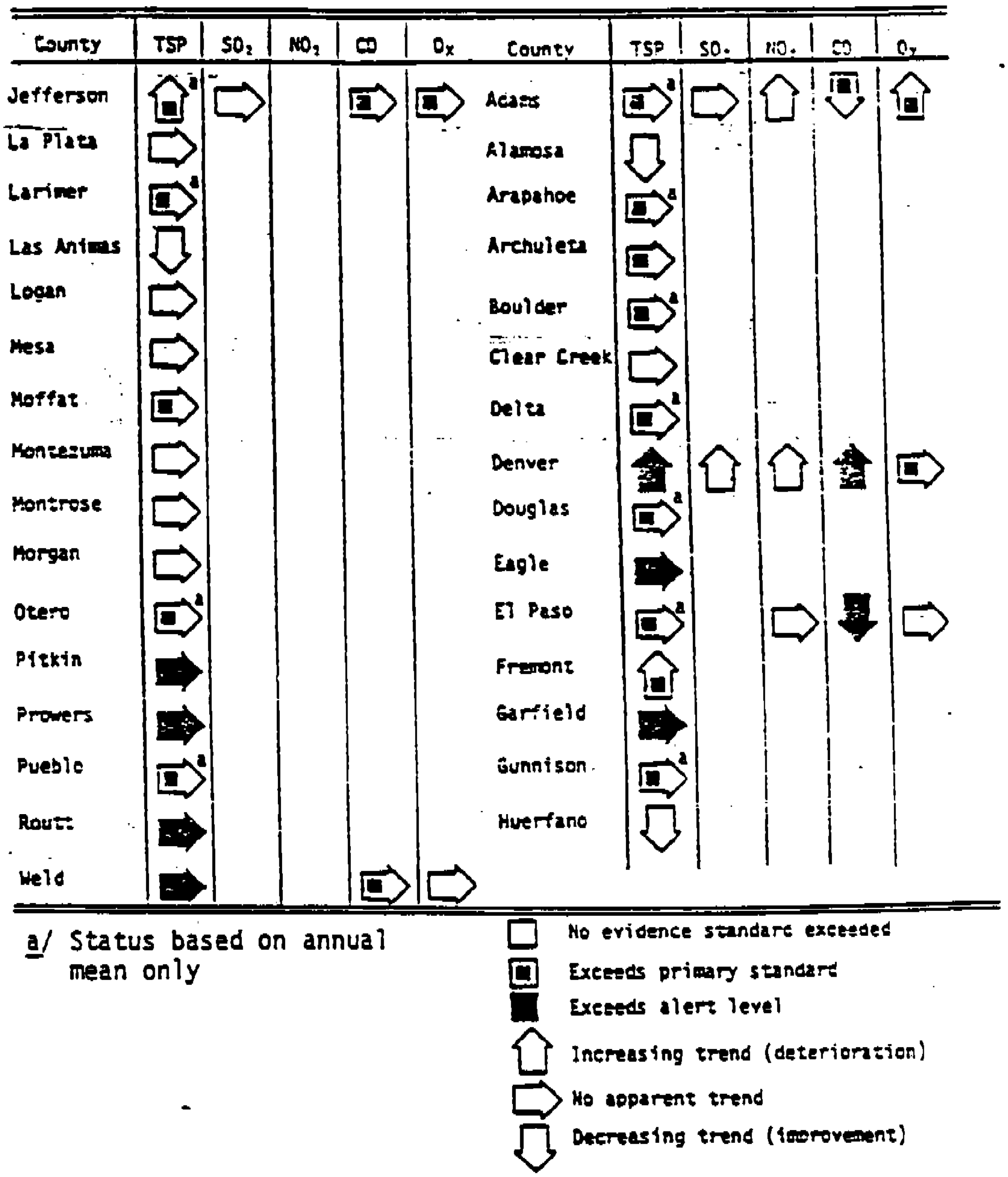


Figure 8. Status and trends in air quality in Colorado.

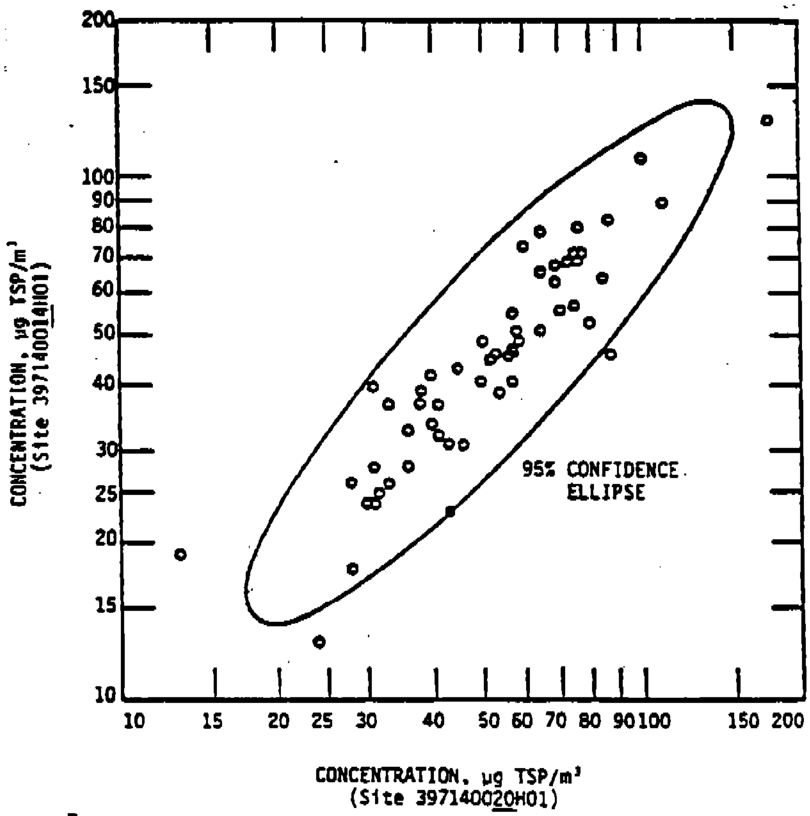


Figure 9. Intersite correlation test data.

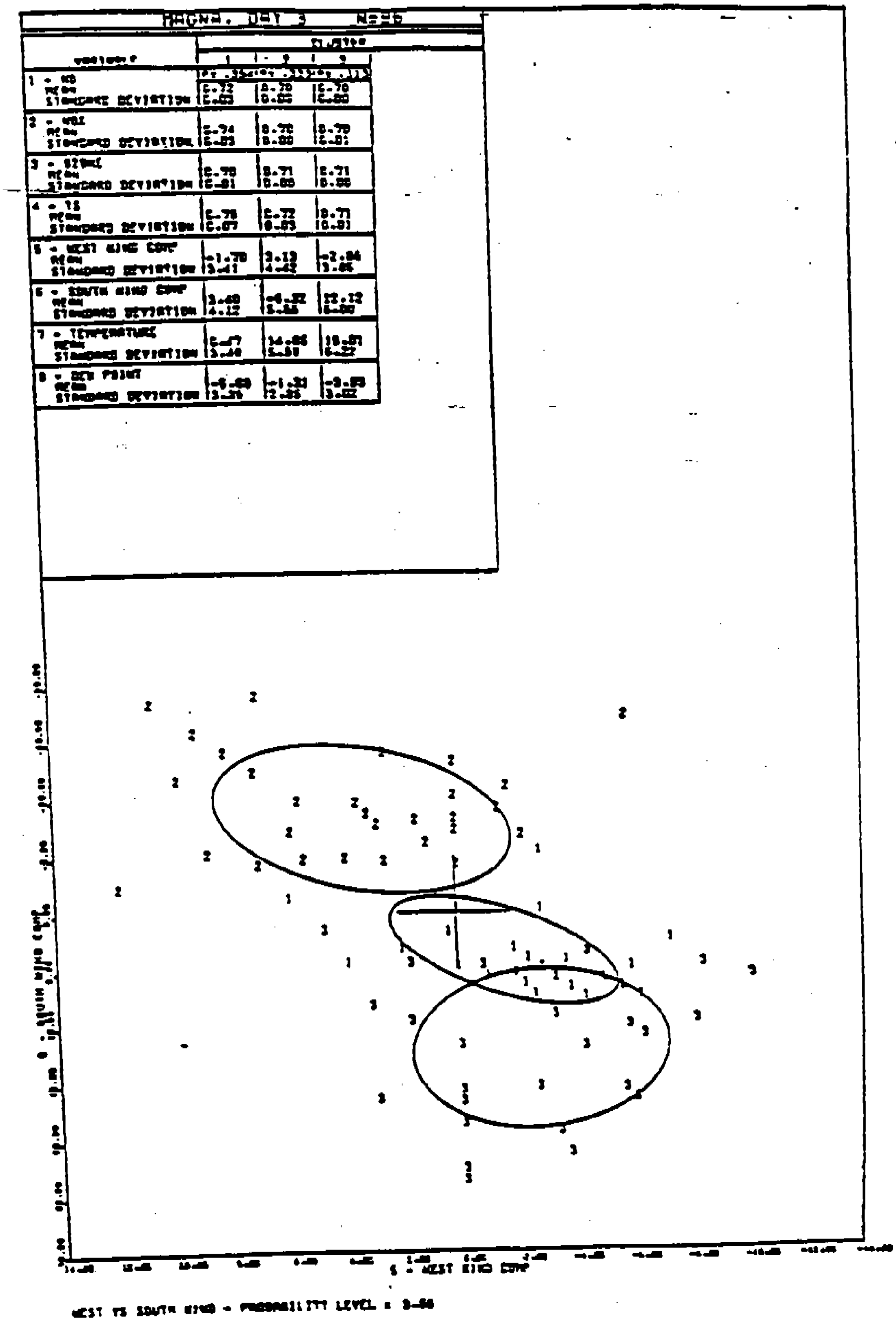


Figure 10. Magna, Utah, day 3, 0.50 probability ellipses of the west-east and south-north wind components for three cluster types. Winds from the west and south are positive.

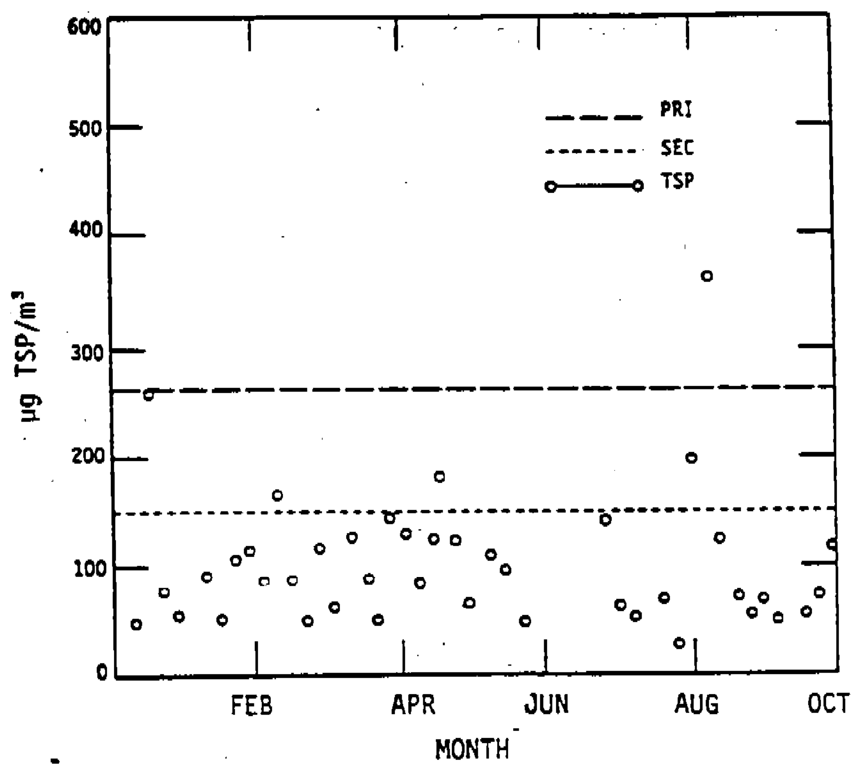


Figure 11. Twenty-four hour TSP values, 1972.

categories (Figure 12).

Point charts usually are accompanied by some "best-fitting" solid line to describe annual trend or diurnal or seasonal pattern. The integrity of the individual points is the foremost reason for displaying data in this way.

8.2.3 Line or Curve Charts

The line or curve chart is perhaps the most widely used method of presenting summarized data graphically. The chart is also the most easily constructed manually. The most common uses in displaying air quality or emissions data are:

1. Data coverage over a long time period (Figure 13).
2. Emphasis on movement rather than on actual amount (Figure 14).
3. Comparison of several series (same measurement unit) on same chart (Figure 15).
4. Trends in frequency distribution; e.g., population exposure (Figure 16).
5. Use of the multiple amount scale (Figure 17).
6. Estimates, forecasts, interpolation, or extrapolation (Figure 18).

8.2.4 Surface Charts

The simple surface or band chart depicts a single trendline with shading or crosshatching filling in the area between trend and base lines to enhance the picture of the trend. Thus, in this chart type:

1. The magnitude of the trend is emphasized,
2. A cumulative series of components of a total trend is depicted, and
3. Certain portions of the chart are accented for a specific purpose.

Figure 19 exemplifies classification of CO data by concentration level and by number of monitored days.

8.2.5 Column/Bar Charts

Column/bar charts are intuitively simple for most readers to follow since they accent discrete dates or categories with comparative heights of the columns/bars for one main statistic. The primary purpose is to depict numerical values of the same type over a given time period (i.e., multiple years), as shown in Figure 20. Other uses are:

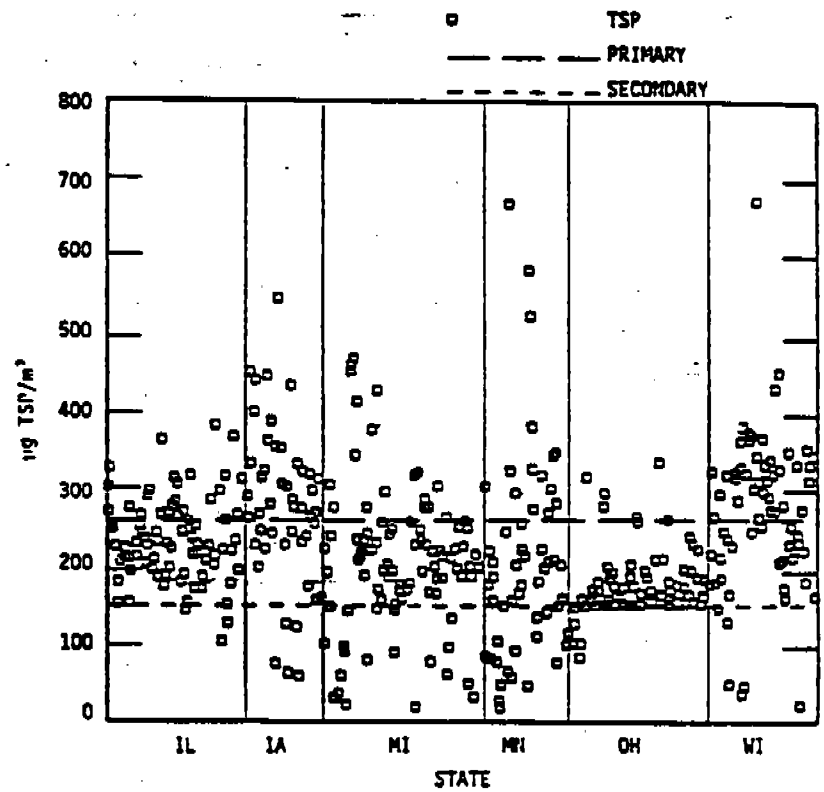


Figure 12. Air quality data, 24-h TSP concentration values, October 15, 1976.

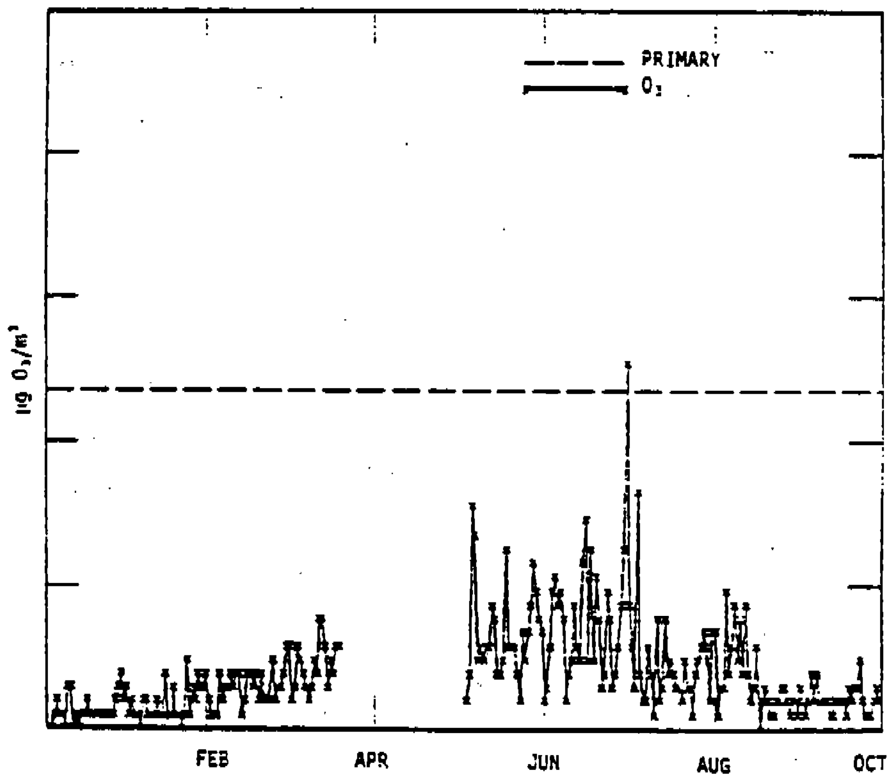


Figure 13. Maximum 1-h O_3 values/day, 1977 (SAROAD site 141220002P10).

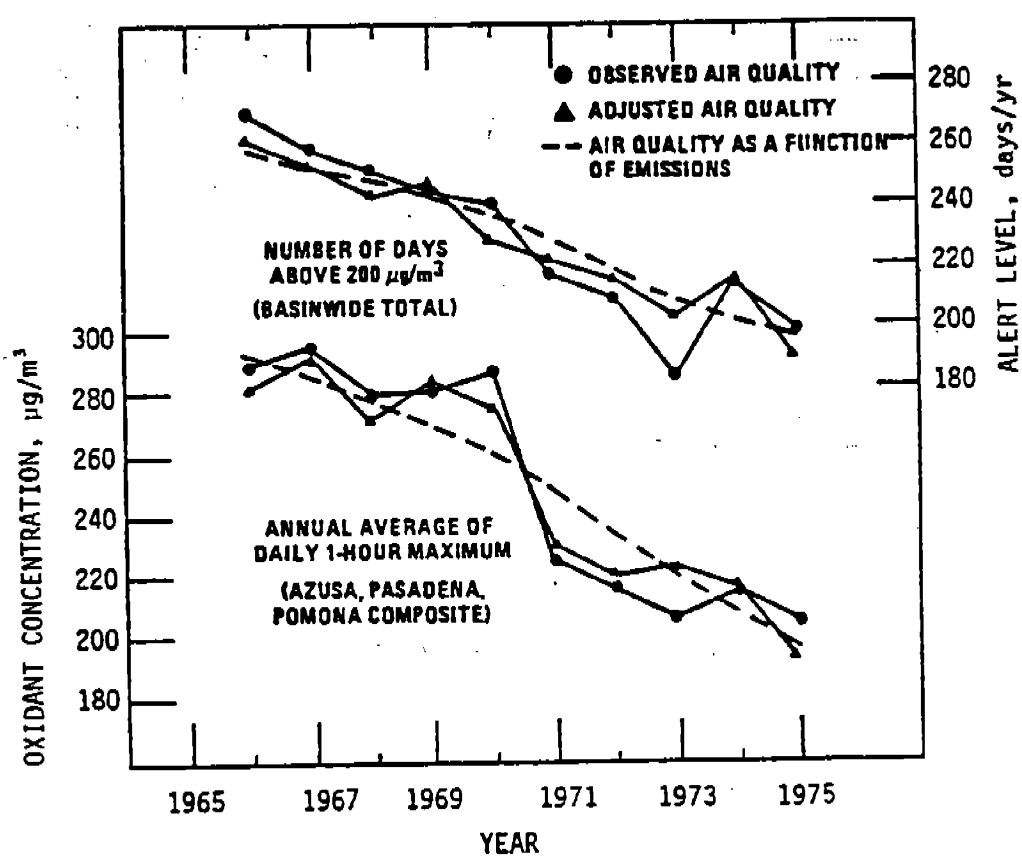


Figure 14. Oxidant trends adjusted for meteorology.

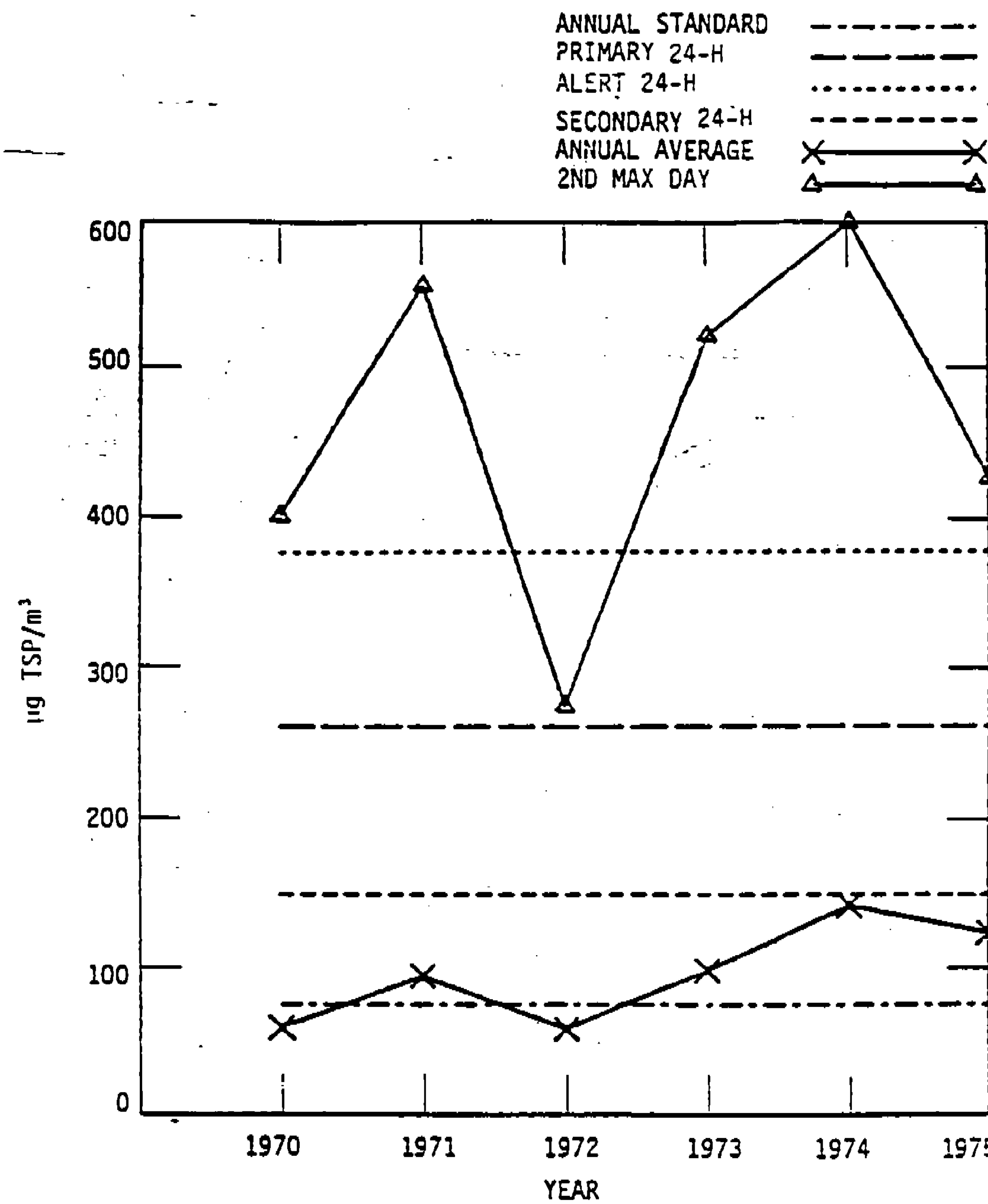


Figure 15. Annual average and second-high day TSP values, 1970-75.

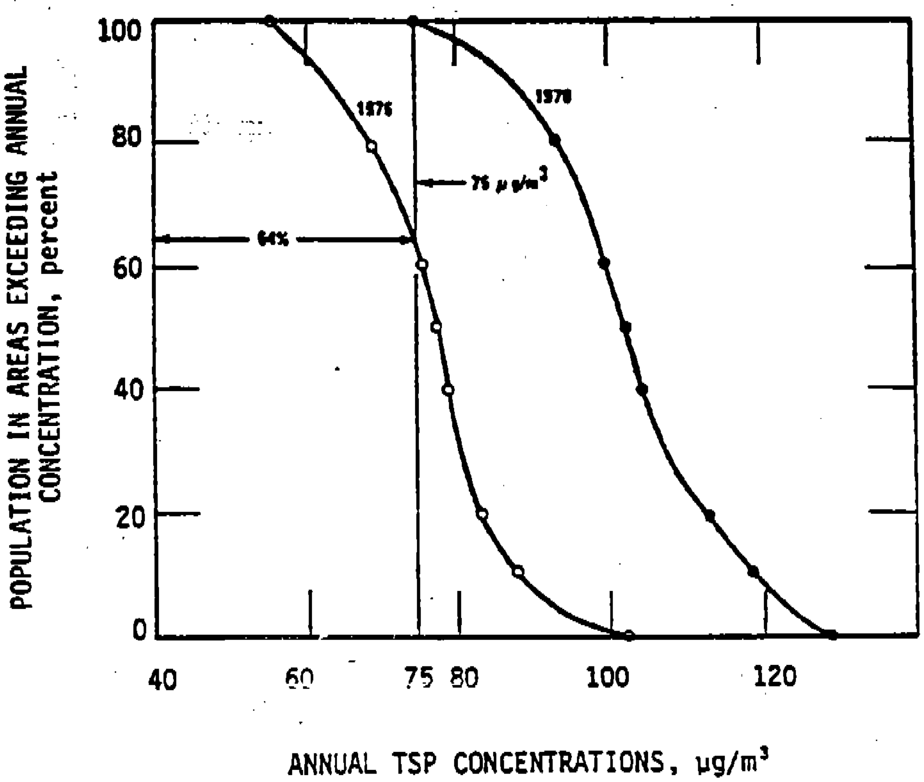


Figure 16. Population exposure distributions of annual mean TSP for 1970 and 1976 in city of Chicago.

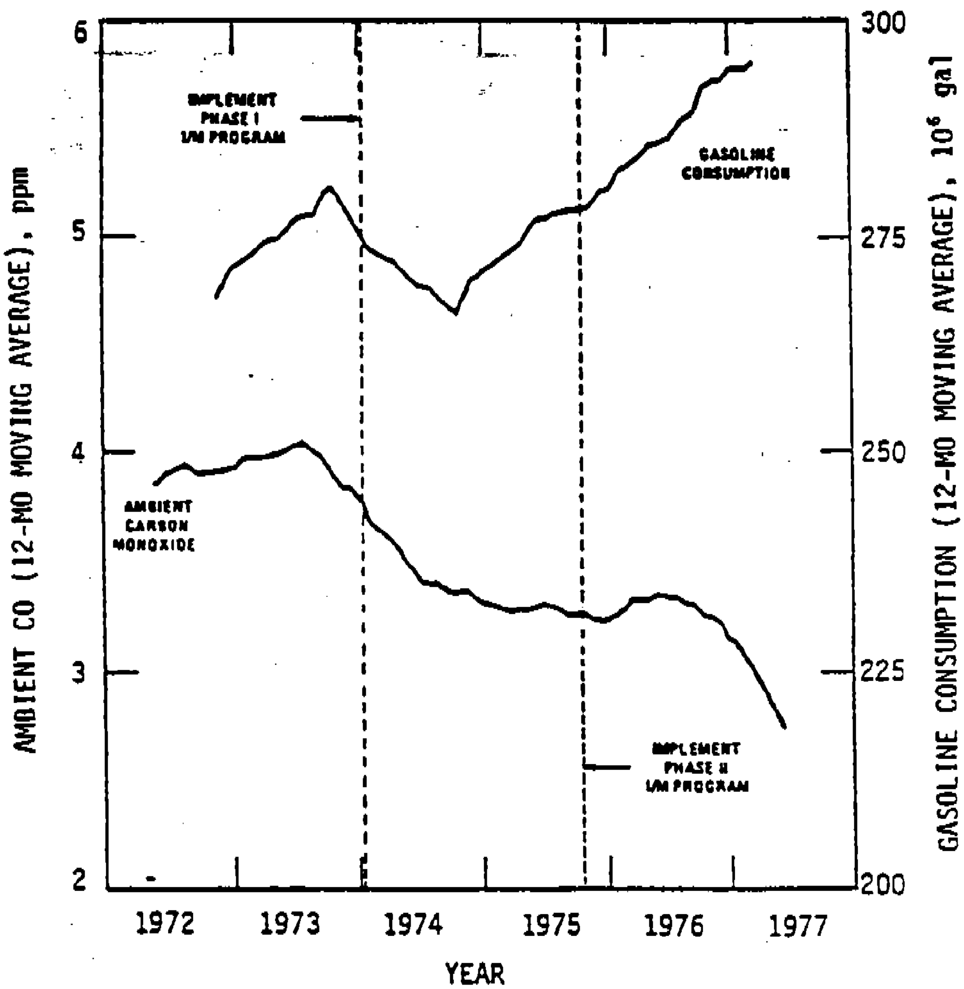


Figure 17. Ambient CO concentration and gasoline consumption, 1972-77.

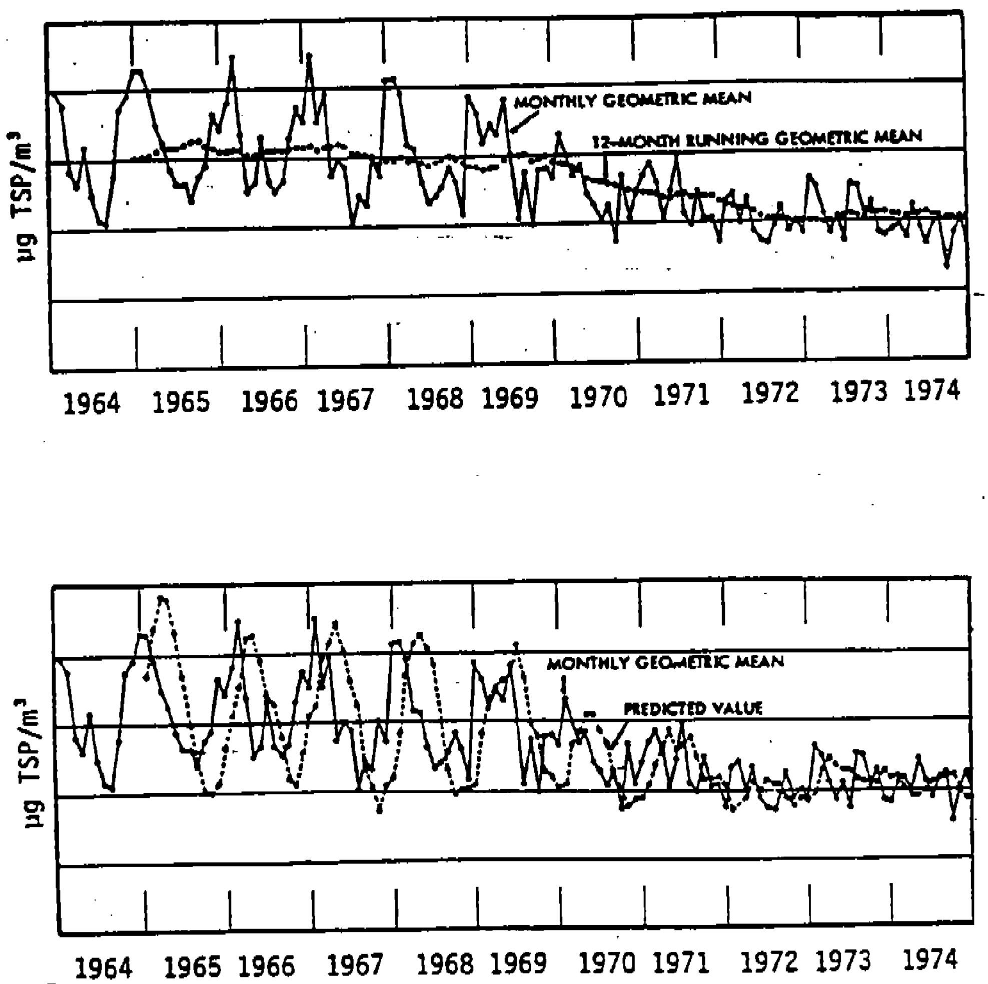


Figure 18. Comparison of monthly GM, 12-mo running GM, and predicted monthly means (by double moving average method), 1964-74.

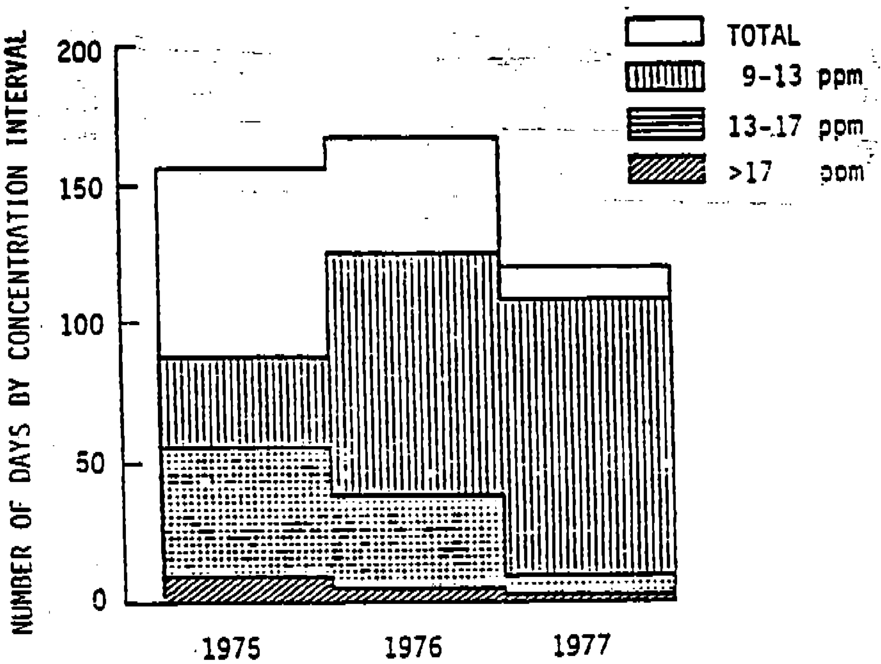


Figure 19. Trends in CO levels in New York's 45th Street Station, April-January, 1975-77.

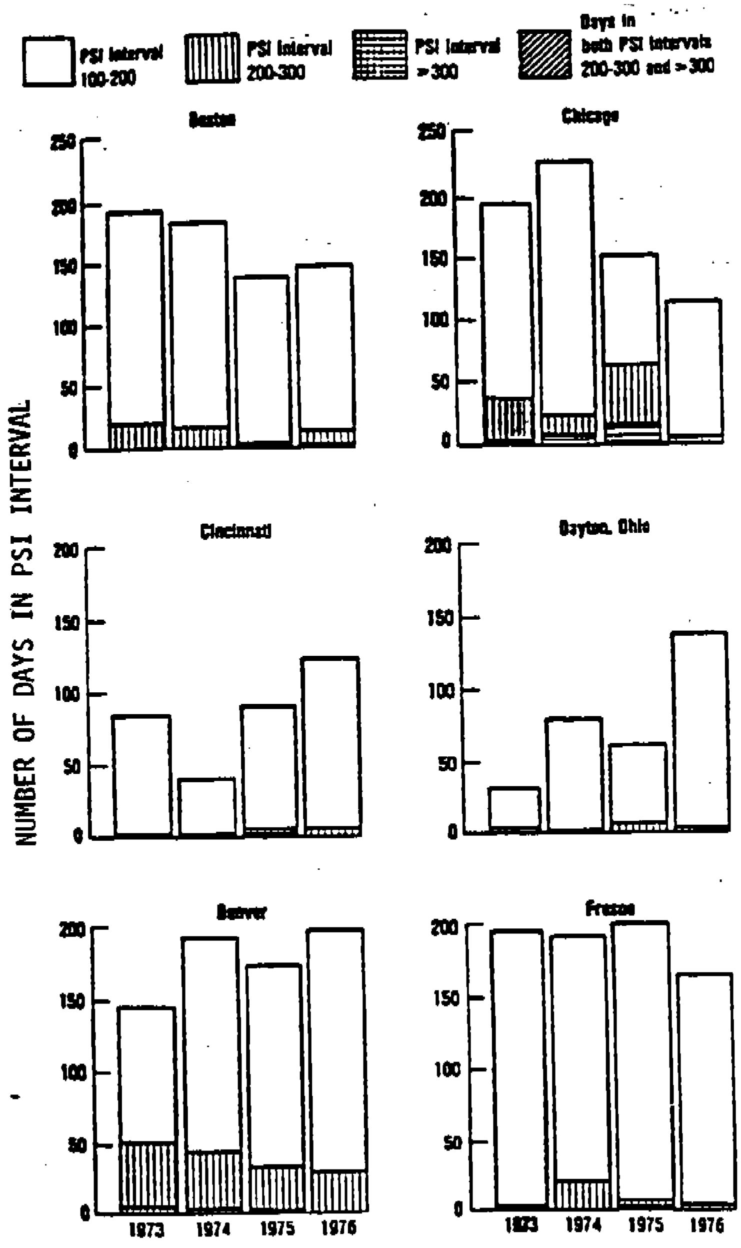


Figure 20. Trends in PSI levels, 16 cities, 1973-76.

1. Comparison of numerical values of the same type for several categories. (Figure 21),

2. Comparison of two or three independent series over time such as grouped columns, subdivided columns, and three dimensions (Figure 22),

3. Display of increases or decreases, losses or gains, or deviation from requirement or norm (Figure 23),

4. Display of ranges of maximal and minimal values for a series (Figure 24).

The wind rose is a familiar application of the bar graph in circular form (Figure 25).

Caveats to be considered in preparation of these charts are irregular time sequences, spacing between columns/bars, scale breaks, shading, and the ordering or sequencing of items represented by the columns/bars.

8.2.6 Pie/Sector Charts

The familar pie/sector chart in the form of a circle compares component parts, and shows their relation to the whole. Source emission categorization lends itself well to this chart type (Figure 26), as does population exposure (Figure 27). The pie chart is often used with line, column, bar, or map displays to exhibit geographic or categorical components of trends.

8.2.7 Map Charts

Map charts are attention-getters, and they are most applicable to environmental statistics, especially air quality. National, regional, State, county, and city maps may depict formation and transport of air pollutants in a real-time, dynamic sense. Moreover, within boundaries, current year-of-record and trends may be shown through isopleths, symbols, or shading. Figures 28 and 29 demonstrate two applications of the treatment of data in the dynamic (isopleth) and static modes.

8.2.8 Three-Dimensional Charts

Advanced computerized graphic techniques have made the three-dimensional chart type a viable alternative to three-way tables and to restrictive two-dimensional plane representations. The most significant applications have been to dispersion modeling and contour mapping (Figure 30).

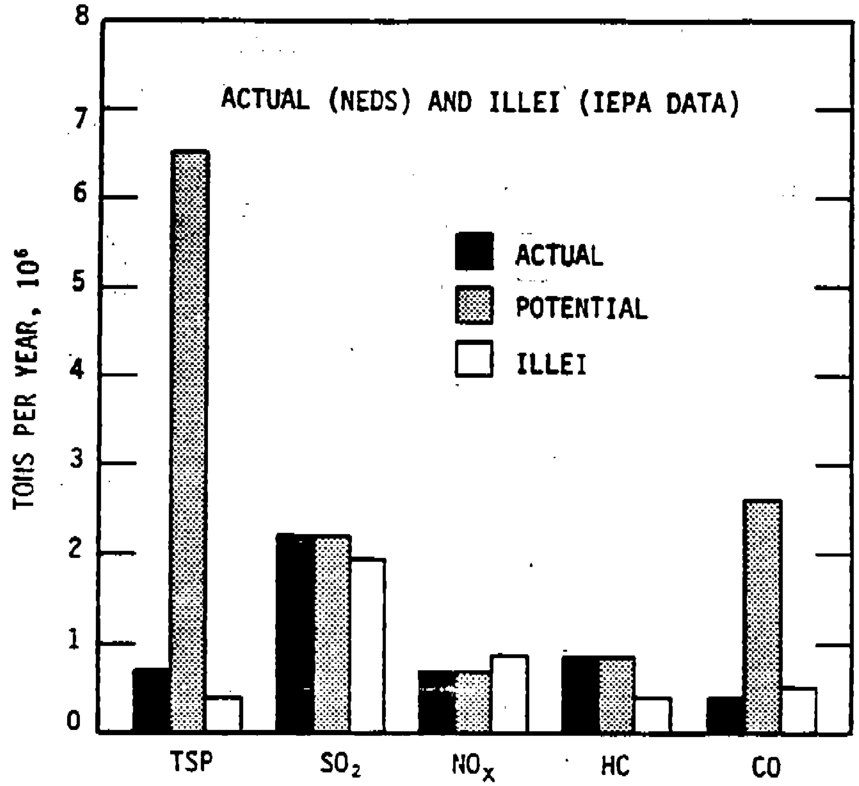


Figure 21. Actual vs. potential emissions for Illinois, tons/year.

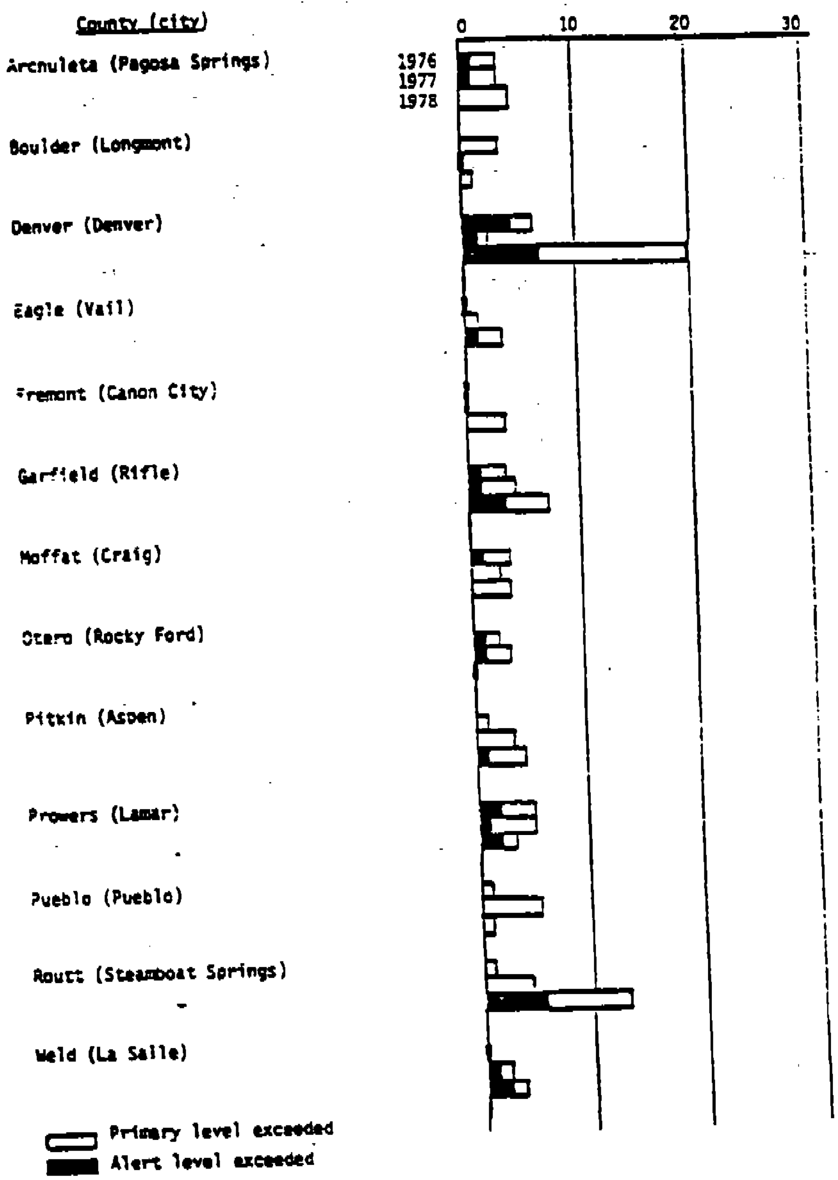


Figure 22. Number of days per year that the TSP primary standard or alert level was exceeded, Colorado.

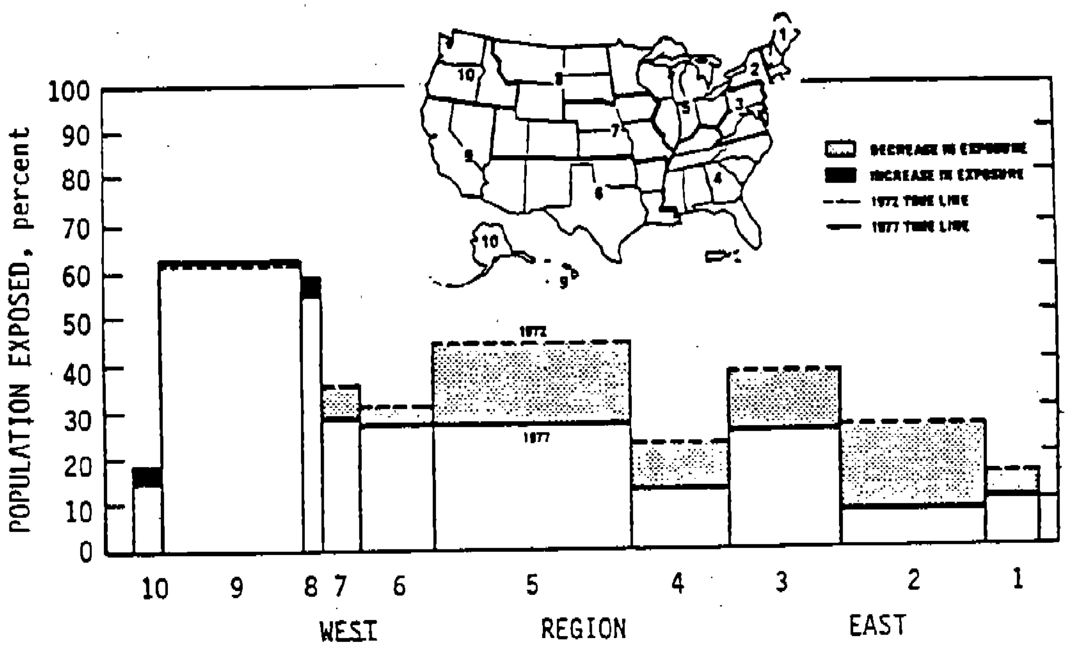


Figure 23. Regional changes in metropolitan population exposures to excess TSP levels, 1972-1977 (width of each regional column is proportional to its metropolitan population).

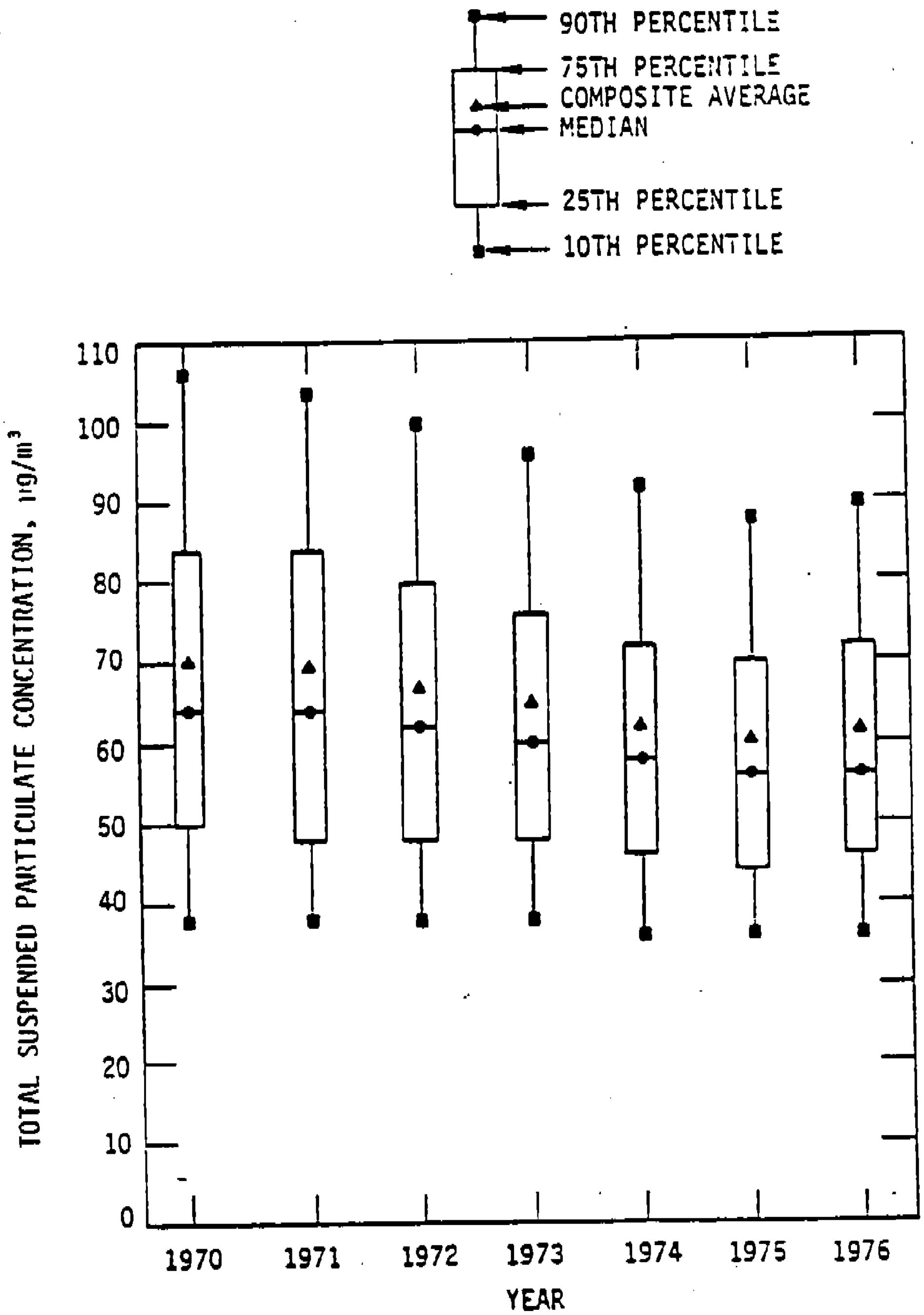
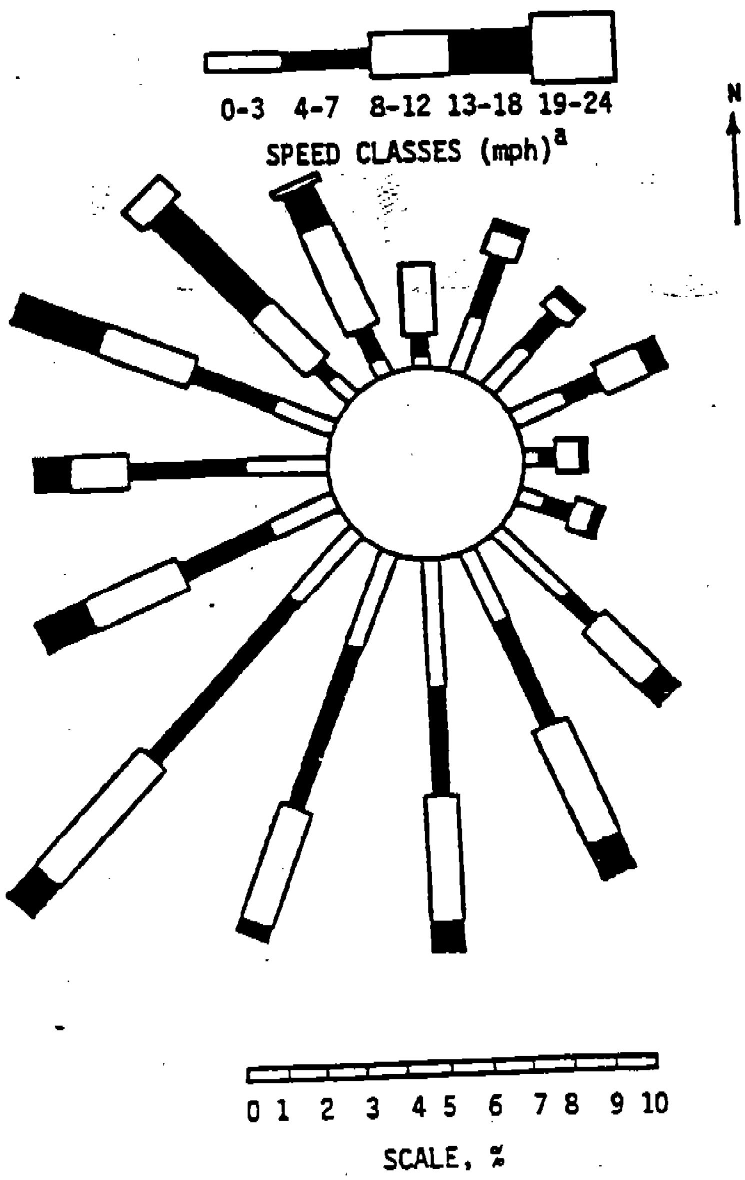


Figure 24. Trends of annual mean TSP concentrations from 1970 to 1976 at 2350 sampling sites.



[a]Bias removed and calms distributed.

Figure 25. Wind rose pattern.

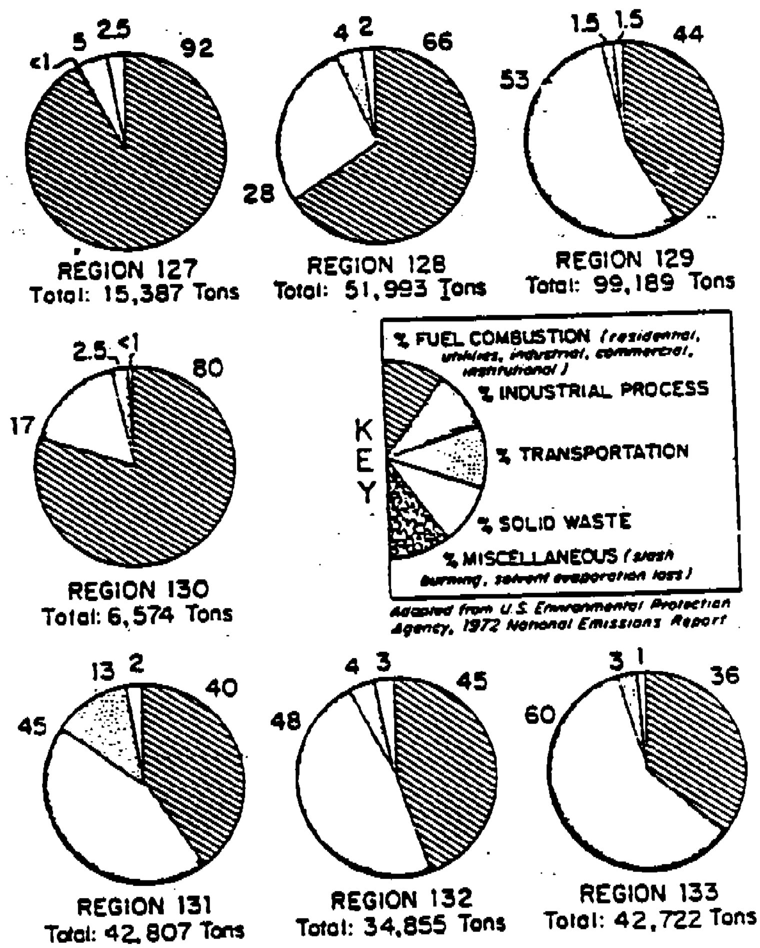


Figure 26. Source category contributions to particulate air pollutants.

TOTAL SUSPENDED PARTICULATES (TSP)

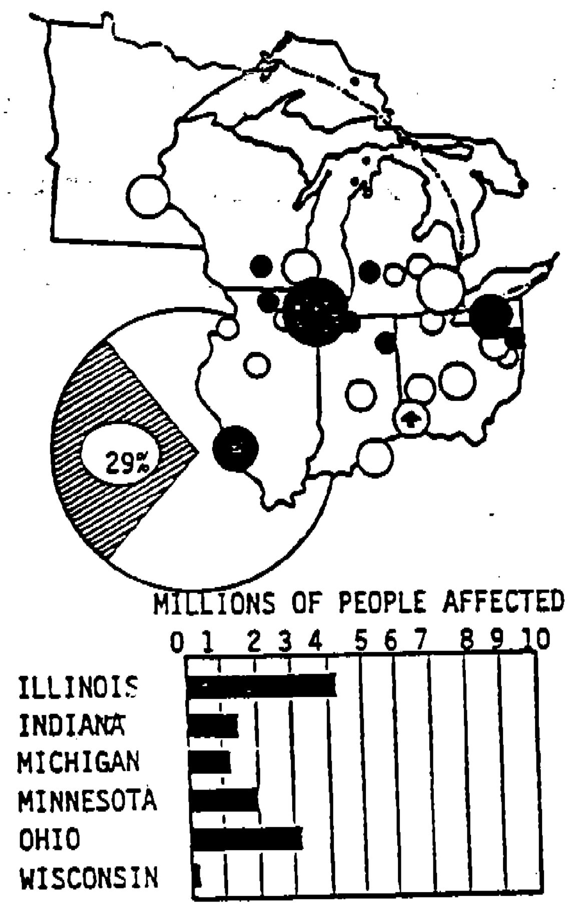


Figure 27. Air quality status (TSP) and trends in 25 largest urban areas in EPA Region 5. Pie chart depicts the percent of population exposed to levels greater than NAAQS for TSP in Region 5. Bar chart is estimated number of people exposed to these exceedances on a state-by-state basis.

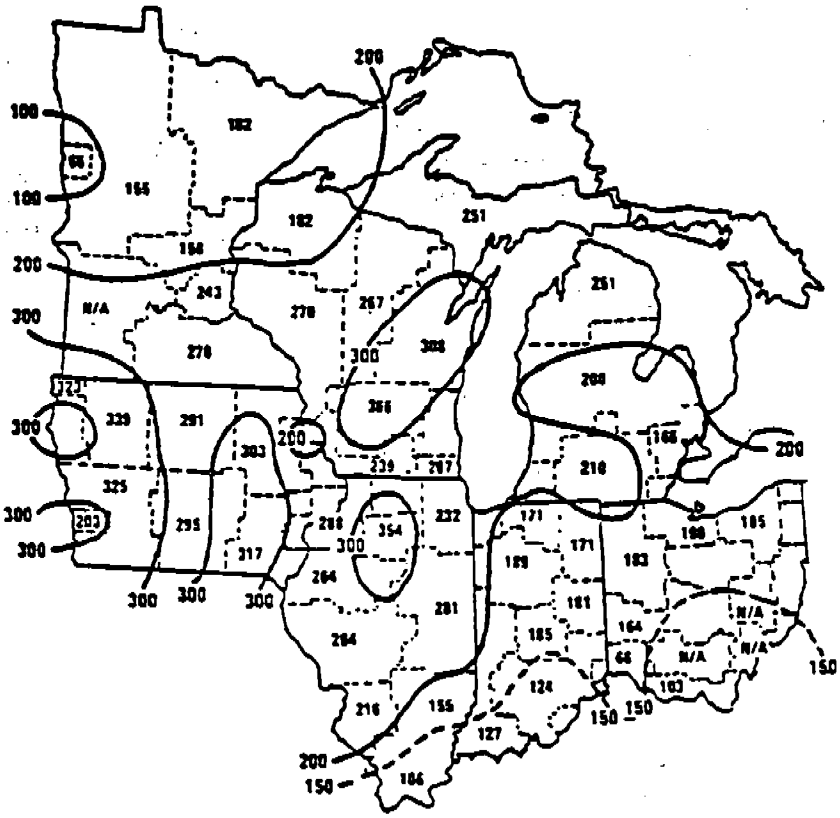
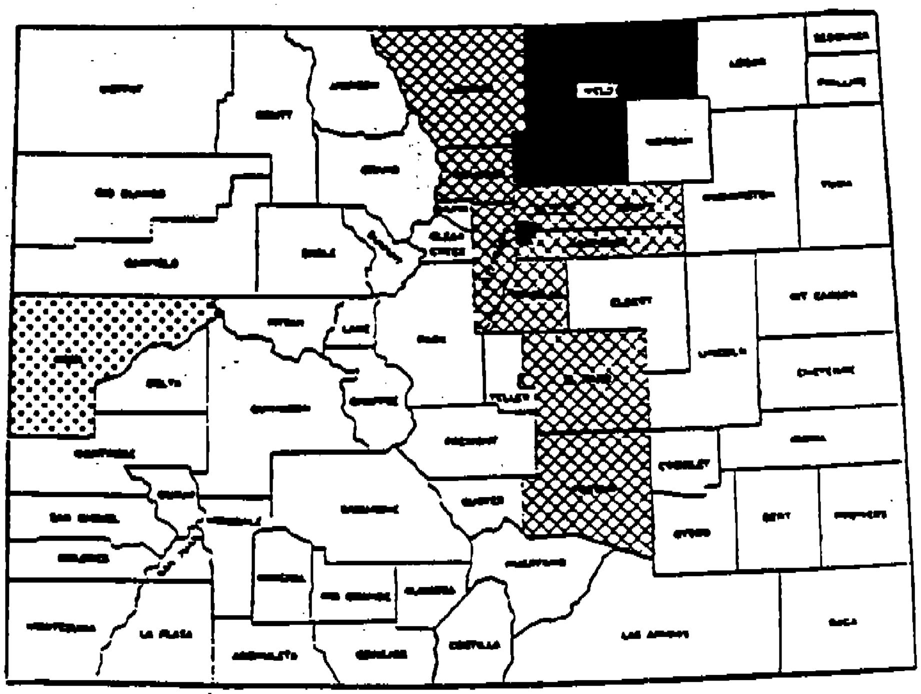


Figure 28. Isopleths of TSP concentrations (μg/m³) in EPA Region V and Iowa for October 15, 1976.



- Insufficient data (<75% of maximum possible observations)
- No evidence primary standard exceeded
- Primary standard exceeded
- Alert standard exceeded

Figure 29. Air quality status, Colorado, 1972.

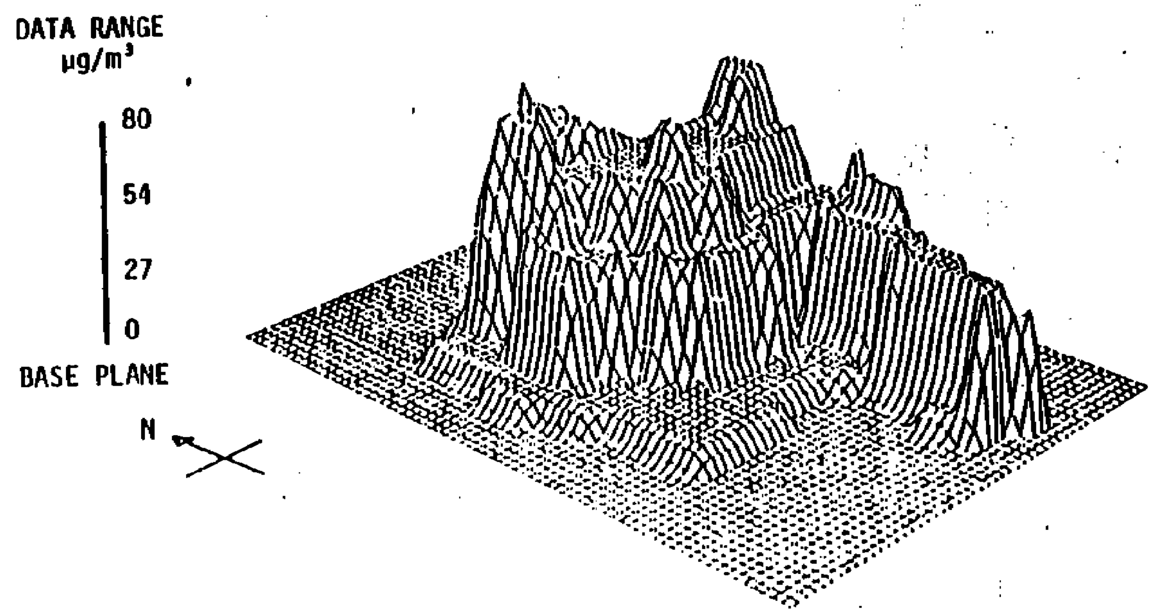


Figure 30. Three dimensional plot of NO_2 concentrations, µg/m³, November 1973.

8.3 CLASSIFICATION OF DATA

One of the early goals for standardizing the codes to be applied nationally to air quality monitoring sites was selectivity for analysis. The SAROAD codes provide time classification, general geographic classification, and site/neighborhood classification. However, only the first two are currently available at NADB for user applications. A profile of environmental quality on any geographic basis will require more unique ways of data classification and analysis in the future.

8.3.1 Time Categories

Overall, the NAAQS's dictate the following categories as most important: year (TSP, SO_2, NO_2), quarter (Pb), day (TSP, SO_2), 8 hours (CO), 3 hours (SO_2), and 1 hour (O_3, CO). However, trends analysis techniques could require classification by hour of the day; day of the week; week; and month for multiple year comparisons. Autocorrelation techniques may require other nonstandard time intervals and hence a more flexible data base retrieval system. Research needs will require further analysis of long- and short-range transport, and time selection will be critical for trajectory analysis.

8.3.2 Geographic Categories

Summary statistics are usually retrieved from NADB on a monitoring or site point-source basis. All individual sites may be retrieved for a city, county, Air Quality Control Region (AQCR), and State for a given time span. Aggregation statistics are not yet standard selection options, except in special computer routines not available to all users. Nationwide reports either summarize individual site statistics for all States or select certain geographic regions, urban areas, or special interstate areas. Likewise, regional and state reports aggregate individual site statistics from all sites or from selected counties or cities. Multicounty, county, and city agencies have more opportunity to deal with specific subcounty areas in annual quality reports.

8.3.3 Site/Neighborhood Categories

This largely unexplored categorization has the potential to be a most interesting summary in future regional and national reports. It is now a part of the National and Regional Environmental Profile (NREP) effort. The new monitoring regulations require that sites be classified according to neighborhood and monitoring objective. Site types published previously were industrial, commercial, residential, and agricultural within center city, suburban, near urban, rural, and remote areas. The new classifications are microscale, middle scale, neighborhood scale, and urban scale. Once established, NAMS will be categorized and summarized in any of the classification schemes.

8.4 INPUT PARAMETERS, DATA TRANSFORMATIONS, AND STATISTICAL COMPARISONS

Table 14 outlines the details of data types which could be summarized. Note that effort was made to delineate individual air quality and emissions statistics from aggregate statistics and counting summaries. Counting summaries are generally used internally as management tools. Only a few such summaries should be related in RNEP's or in reports to the Administrator. Much work remains on standardizing data transformation techniques. In addition, the techniques of maximum-likelihood estimation, application of the general linear model (GLM) regression and analyses of variance, and spectral (time-series) analysis could be applied in more cases than they are presently.

8.5 AUDIENCE APPLICABILITY

The most important consideration in data presentation is applicability to a specific audience. For example, the report to the Administrator may cover not only air quality assessment but also the end results of EPA programs and regulatory policies. The data needs are national in scope for the environmental assessment, perhaps with highlighted urban areas on criteria and noncriteria pollutant issues. Both air quality and emissions data, current status and trends, would be given on a national and possibly a region basis; this level would require more graphic presentation than a technical/management report--for example, a graphical summarization of data resulting from singular national events (volcanic eruptions or widespread duststorms).

A report to the Regional Administrator may be merely a subset of the report to the national Administrator, but with emphasis on regional policies. The summary statistics are usually given state by state, by selected major urban areas and by regions.

Reports to other government agencies and in response to congressional and (to a large extent) public inquiries are usually limited to specific areas throughout a State. Environmental profiles provide an easily understandable

TABLE 14. OUTLINE OF INPUT PARAMETERS, DATA TRANSFORMATIONS, AND STATISTICAL COMPARISONS

- Individual ("raw") air quality data
  - All hourly concentration values
  - All daily concentration or "index" values
  - Maximum/minimum site values
  - Average/median site values
  - Quantiles by site
- Individual emissions data
  - Process
  - Stack
  - Plant
- Aggregate air quality data for given category
  - Maximum/minimum values
  - Average or median of maximum/minimum data values
  - Weighted average/median of averages/medians
  - Maximum/minimum quantiles (e.g., max all 90th percentiles)
  - Average/median quantile of like quantiles
- Aggregate emissions data for given category
  - Maximum/minimum process
  - Maximum/minimum stack
  - Maximum/minimum plant
  - Average/median process
  - Average/median stack
  - Average/median plant
- Counting summaries for given statistic
  - Number of sites, cities, counties, States, regions
  - Number of processes, stacks, plants, cities, etc.
- Data transformations
  - Distributional (e.g., Weibull, logarithmic)
  - Regression
  - Analysis of variance (ANOVA)
  - Spectral analysis and smoothing techniques
- Example statistical comparisons by site, geographic, and time categories
  - Total number of pollutant sites
  - Number of sites exceeding annual standard
  - Number of sites exceeding short-term standards
  - Counties with data
  - Counties exceeding standards
  - Days exceeding primary standard
  - PSI class distribution
  - Site averages by year
  - City averages by year
  - Counties (urban areas, etc.) with significant pollutant trends up, down, no change
  - Emission estimates by major source category
  - Emission density vs. population and land use

graphic summary of environmental status and trends, and solve the problem of repetitive requests.

Air quality standards "violations" and trends on a pollutant/county or urban area basis (using colored maps) are the main tools of the RNEP effort. The most common question--Where are the least polluted areas of the United States?--has not been addressed directly in EPA reports because of the complexity of issues regarding sufficiency, representativeness, and comparability of data. However, profiles with an added impetus to incorporate the PSI will at least indirectly address this issue.

Scientific and technical audiences may have stringent requirements for level of data summary and documentation of the statistical analysis techniques.

8.6 CAVEATS AND SUGGESTIONS

The following should be considered in planning graphics displays:

1. Descriptive or statistical parameters required?
2. Time to develop a summary from individual data?
3. Cost for preparation of graphics?
4. Data quality assured, valid, and carefully analyzed?
5. Proper emphasis on purpose of chart?
6. Chart not too light, heavy, confusing, complex?
7. Color/shading required?
8. Lettering styling and size appropriate?
9. Scale heights distorted?
10. Data gaps in chart treated correctly?
11. Too much gridding and/or lettering?
12. Grid units appropriate?
13. Grid points clearly identified?
14. Use of pictorial symbols and descriptive titles?
15. Chart self-explanatory?

8.7 AVAILABLE PLOTTING RESOURCES

Graphic display packages are available at both the USEPA National Computer Center (NCC),[2] Research Triangle Park, N.C., and the Washington Computer Center (WCC) Washington, D.C. "Stand-alone" systems include *Integrated Plotting Package (IPP), Harvard Graphic, Tektronix, Calcomp,* and *Statistical Analysis System (SAS).* Subroutines are available for user program interface under most of these systems. Currently, statistical analysis and graphics programs developed in some EPA Regional Offices are being documented for use by other regions; these include PSI, box-plot, and pollution/wind roses.

8.8 GUIDANCE FOR SELECTION OF CHARTS

No single standardized list of charts should be mandatory for displaying aerometric data; however, the following steps should be used to guide the selection of the most relevant charts.

1. Will the audience be interested in technical details and require follow-up documentation concerning both data base and data analysis methodology? If so, individual data or tabular summary may be sufficient. If graphics are needed, see steps 2-12 below.

2. Select the geographic scale of the summary: national, regional, AQCR, SMSA, county, urban area, city, township, or monitoring site/source.

3. Select concept to be charted: trend, current status, parameter(s) vs. parameter(s), one parameter vs. several categories, composition of components, or maps.

4. Select time class: hour of day; day of week; month; calendar quarter; season; year; or multiple year.

5. Select period of analysis: start year/month/day/hour and end year/month/day/hour.

6. Select statistical summaries: individual, summation, or aggregation.

7. Select statistical analysis technique: descriptive statistics; distributional; regression; ANOVA; spectral analysis; or trend analysis.

8. Select site type, if desired: industrial, commercial, residential, and so forth.

9. How can the analysis be accomplished? Manually (small data base), hand calculator/computer, or large computer.

10. If a computer is necessary, are there statistical analysis and/or graphics systems available? Check NCC, WCC, and Regional Office.

11. Retrieve data according to selection criteria above.

12. In preparation of the chart, follow Section 8.7 caveats.

8.9 SUMMARY AND RECOMMENDATIONS

This section discussed established data display techniques applicable to aerometric measurements, and provided guidance for determining the formats most suitable for the concepts to be displayed, the audience applicability, the input data classification and analysis, and the plotting resources. The discussion was limited to data resulting from the air monitoring and emissions inventory processes. The purpose of such data is to provide quantitative insights to help abate air pollution, manage natural resources, plan environmental programs, and inform the public.

The discussion did not consider topics such as sufficiency, representativeness, and data validity, and it assumed data to be efficiently accessible. Since new statistical analysis techniques are covered in preceding sections, this section discusses only analysis by currently available techniques.

Finally, the section did not attempt to standardize graphic formats; it merely provided guidance and criteria, and gave example displays of suitable data presentations. Graphics must accurately portray the data, help readers understand the data, and help captivate interest. The guidance and criteria in this section are not meant to stifle the innovation and creativity of those preparing these artforms.

8.10 FUTURE ISSUES

Section 8 will be modified as new statistical techniques are used to yield different parametric relationships and thus different graphics. Mandatory standardized analysis techniques may result in standardized computer graphics programs. Questions to be addressed in the future are:

1. How stringent will USEPA be on data completeness and "representativeness"? Should we display quality assurance data? If so, should we exclude source-impacted monitoring sites, or restrict analysis to NAMS?

2. How can special "success stories," nonattainment status, or other policy/regulatory issues be displayed in environmental assessment documents? Should they be included?

3. Will population exposure analysis be used in future reports to the Administrator? Will population exposure software be easily used by Regional Offices?

4. How frequently will the RNEP's be published? How do the profiles relate to the report to the Administrator, to Congress, and to the *National Air Quality and Emissions Trends Report*? Should the profile format change from year to year and perhaps cover selected urban areas or regions in "off" years?

8.11 REFERENCES

1. Spear, Mary E. Practical Charting Techniques. McGraw-Hill, Inc., New York, N.Y., 1969.

2. NCC User Reference Manual.

9.0 CONTINUITY OF YEAR-TO-YEAR REPORTS

This section focuses on two major topics which can seriously affect the continuity of data reported throughout time. First are changes in the operational definitions for measurements of pollutants and in the statistical indicators or techniques. Second are shifts in the data base.

The discussion assumes that detailed data, as opposed to summary values, will continue to be available from NAMS. It also assumes that as data needs change with time, published reports will reflect the new priorities rather than maintain format merely for historical continuity.

9.1 METHOD CHANGES

Data analysts preparing air quality reports should continually check for changes in measurement techniques. These changes may require the adjustment of the raw data for past years to maintain consistency in the data base.

The data base should not prevent the adoption of a new statistical indicator (for example, a change from the median to the mean as an indicator of the central tendency of a distribution). Availability of detailed data for past years should make it possible to compute values of the new indicator for data of the past. There should be no reason for presenting trend tables with old indicator values for past years and new indicator values for the present year. Similarly, introduction of a different statistical method should be accompanied by a presentation of its applicability to data of the past.

9.2 NAMS NETWORK CHANGES

The discussion above assumes that the number and location of monitors providing data is held constant over the years for which the trend is studied. However, the NAMS network will probably undergo many changes in years to come as more is learned about the nature and effects of the various pollutants, as funding levels change, and as source configurations are altered. If summarization of data is carried out on the urban level (as recommended in Section 4 instead of by monitoring site, effects of changes in the network will be minimized. If an index of air quality is adopted, it is recommended that the subset of NAMS used for index determination remain unchanged over a relatively long period.

There are a number of ways in which changes in the NAMS network could occur; each would have a distinct effect on the continuity across years. In particular, there are three changes that could occur even if the subset of NAMS used for index determination remains fixed.

1. The stations could be resited within an urban area,
2. The network in an urban area could be expanded, or
3. The network in an urban area could be contracted.

Effects of these changes can be minimized by the procedures given below.

A station should be resited only if there is hard evidence to justify it. If resiting occurs, data from both the old and new sites should be gathered concurrently for one year to firmly establish the differences between the two sites and to assist in correcting any misleading conclusions formed using data from the old site.

A change in location for a station used to report the maximum concentration for an area should not affect the continuity of the data base; in fact, the maximum monitoring site should be reviewed periodically to assure a reasonable approximation of the maximum concentration associated with the urban area.

It is difficult to foresee plausible reasons for resiting those stations located according to the criteria of high population density and poor air quality, unless there is a considerable change in air quality patterns over the urban area. In this case, the resulting "discontinuity" of the data base would reflect a discontinuity in the measured phenomenon and should be noted.

If the NAMS network in an urban area is expanded, past values of indices and trend analyses can be adjusted by relating concurrent values from the old network stations to those from the new combined network.

If the NAMS network in an urban area is contracted, all indices and trend computations for past years should be redone using only the data from the smaller network.

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-450-4/81-015	2.	3. RECIPIENT'S ACCESSION NO. February 1981
4. TITLE AND SUBTITLE U.S. Environmental Protection Agency Intra-Agency Task Force Report on Air Quality Indicators		5. REPORT DATE 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) W.F. Hunt, Jr., (Chairman), G. Akland, W. Cox, T. Curran, N. Frank, S. Goranson, P. Ross, H. Sauls and J. Suggs		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Environmental Protection Agency Offices of: Air, Noise and Radiation, Research and Development and Planning and Management EPA Region 5		10. PROGRAM ELEMENT NO. 11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS U.S. Environmental Protection Agency Office of Air, Noise and Radiation Office of Air Quality Planning and Standards Research Triangle Park, N.C. 27711		13. TYPE OF REPORT AND PERIOD COVERED Task Force Report 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

The Intra-Agency Task Force on Air Quality Indicators was established to recommend standardized air quality indicators and statistical methodologies for presenting air quality status and trends in national publications. This report summarizes the recommendations of the Task Force grouped into four categories: data base, data analysis, data interpretation and data presentation. The report includes the position papers prepared by the Task Force members dealing with precision and accuracy data, detecting and removing outliers, area of coverage and representativeness, data completeness and historical continuity, statistical indicators and trend techniques, inference and conclusion, data presentation, and continuity of year-to-year reports.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Intra-Agency Task Force on Air Quality Indicators Air Quality Indicators Statistical Methodologies Data Base Data Analysis Data Interpretation Data Presentation Precision and Accuracy Data Outliers Representativeness Data Completeness Statistical Indicators Trend Techniques Continuity of year-to-year reports		

18. DISTRIBUTION STATEMENT Release Unlimited	19. SECURITY CLASS *(This Report)* Unclassified	21. NO. OF PAGES 105
	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE

EPA Form 2220-1 (9-73)

United States
Environmental Protection
Agency

Office of Air, Noise, and Radiation
Office of Air Quality Planning and Standards
Research Triangle Park NC 27711

Official Business
Penalty for Private Use
$300

Publication No. EPA-450/4-81-016

Postage and
Fees Paid
Environmental
Protection
Agency
EPA 335

If your address is incorrect, please change on the above label; tear off, and return to the above address.
If you do not desire to continue receiving this technical report series, CHECK HERE ☐ , tear off label, and return it to the above address.

APPENDIX K

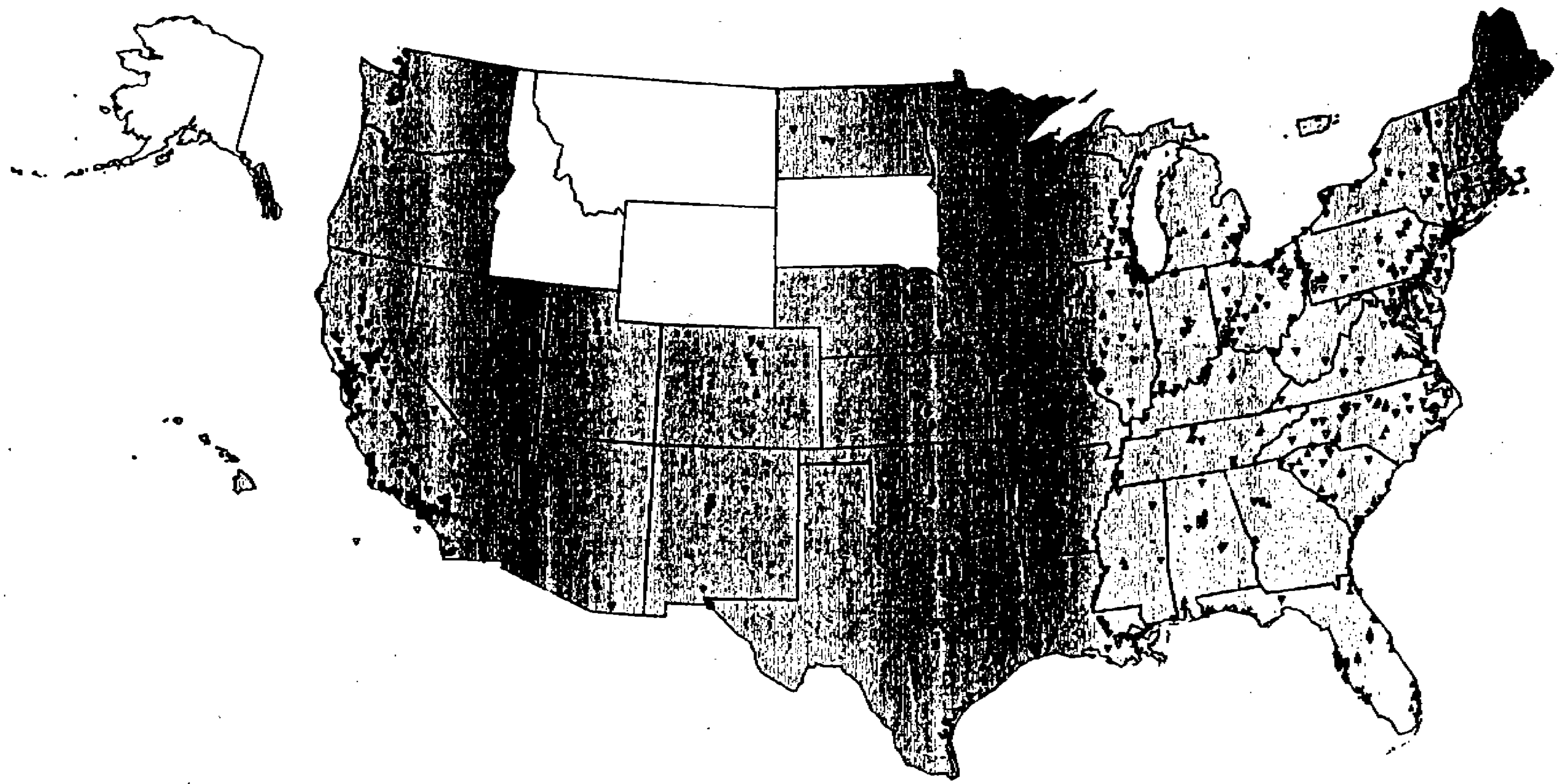
SAMPLE OUTPUTS FROM AIRS GRAPHICS

UNITED STATES

AIR QUALITY MONITORS FOR O[illegible] ONLY)

Monitor Types: NAMS, SLAMS

SHADED STATES HAVE MONITORS

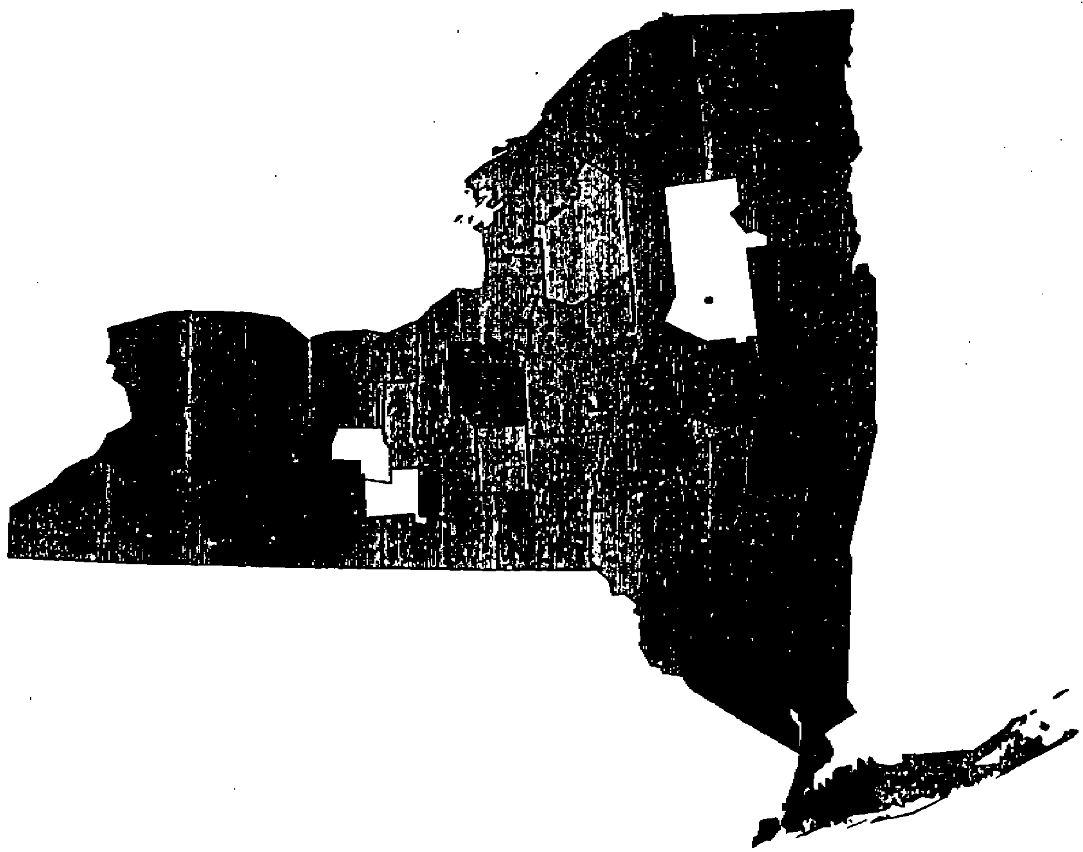


MONITOR LOCATIONS: ▲ (241) NAMS ▽ (579) SLAMS

USEPA, OAQPS, TSD, NATIONAL AIR DATA BRANCH 08/20/93

NEW YORK
AIR QUALITY MONITORS FOR O3 ... NLY)
Monitor Type: (All)

SHADING INDICATES COUNTY POPULATION



COUNTY POPULATION:

- < 25,000
- < 50,000
- < 100,000
- < 300,000
- >= 300,000

MONITOR LOCATIONS:

- ▲ (12) NAMS
- ▽ (30) SLAMS
- ■ (24) OTHER
- | | (0) INDUSTRIAL
- ★ (9) UNKNOWN
- ◆ (0) TRIBAL
- + (0) INDEX SITE
- * (0) NON-EPA

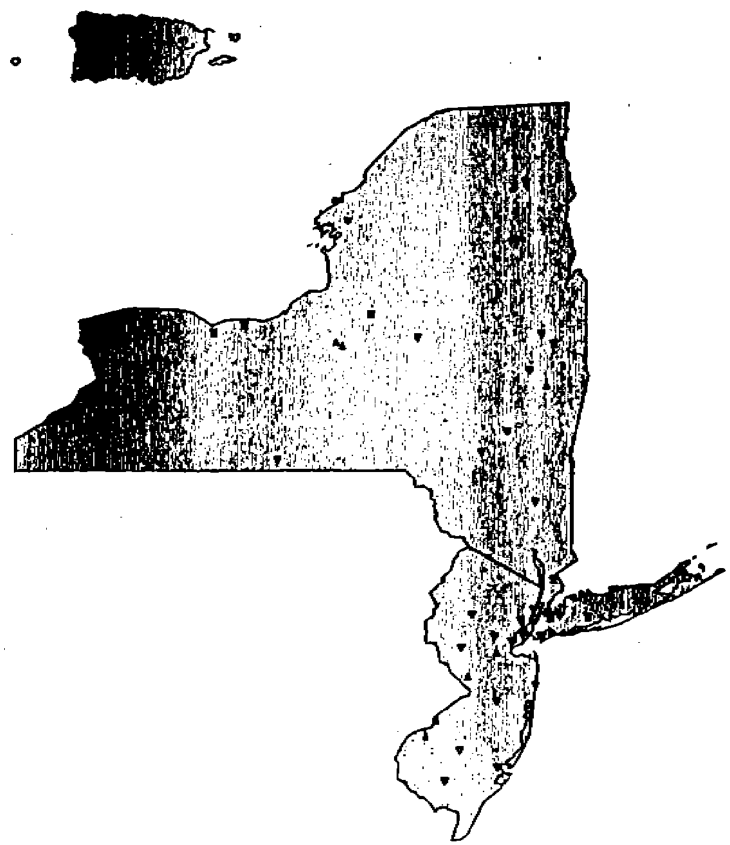
USEPA, OAQPS, TSD, NATIONAL AIR DATA BRANCH

08/23/93

REGION II
AIR QUALITY MONITORS

Monitor Types: NAMS

SHADED STATES HAVE MONITORS



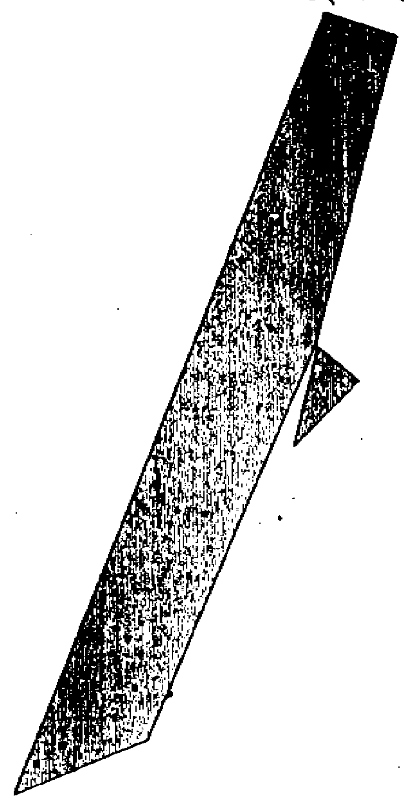
MONITOR LOCATIONS: ▲ (18) NAMS ▽ (42) SLAMS

08/20/93

USEPA, OAQPS, TSD, NATIONAL AIR DATA BRANCH

SHADED COUNTIES HAVE MONITORS

171 TONS (< 1 % OF NATIONAL VOC EMISSIONS)



CIRCLES: 0 MILE RADIUS

MONITOR LOCATIONS:

▲ (0) NAMS ▽ (2) SLAMS ■ (2) OTHER □ (0) INDUSTRIAL

★ (3) UNKNOWN ◆ (0) TRIBAL ✚ (0) INDEX SITE * (0) NON-EPA

■ PLANT LOCATION (2 DISPLAYED)

USEPA, OAQPS, TSD, NATIONAL AIR DATA BRANCH

08/23/93

REGION IV

MONITORS FO

types: NAMS / SLAMS

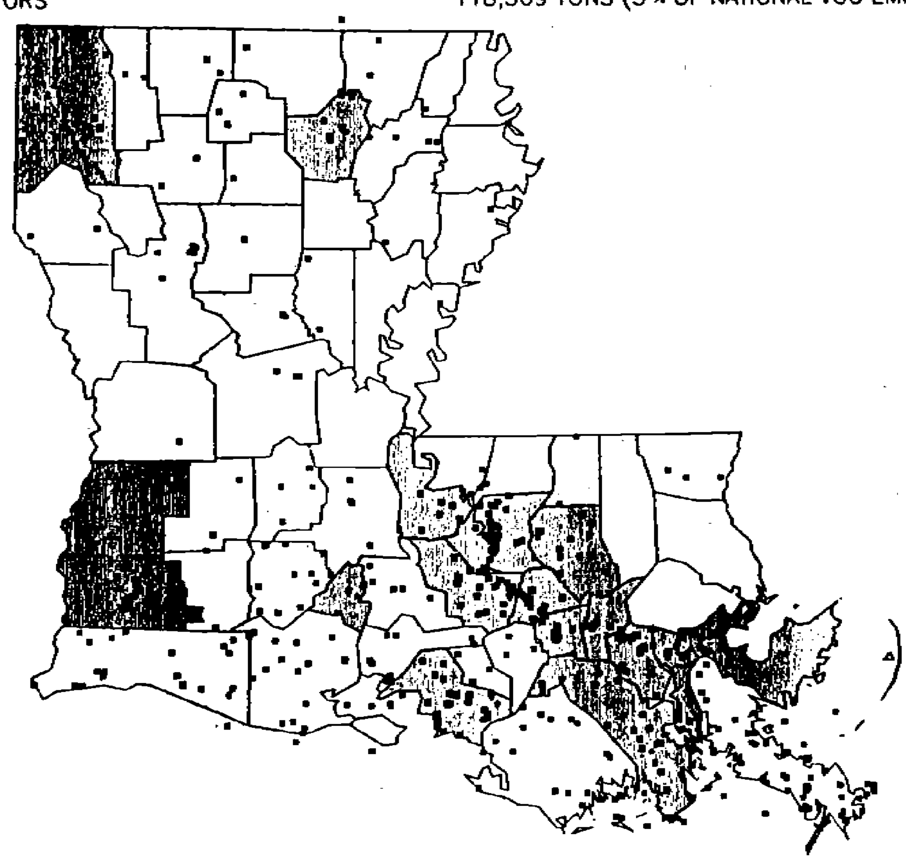
SHADED STATES HAVE MONITORS

MONITOR LOCATIONS: ▲ (62) NAMS ▽ (79) SLAMS

LOUISIANA– AQ MONITORS F[illegible] (ACTIVE ONLY)
(All) – ALL SIC CODES
FACILITIES WITH VOC EMISSIONS [illegible]Y IN 1990 ((All))

SHADED COUNTIES HAVE MONITORS

118,369 TONS (5 % OF NATIONAL VOC EMISSIONS)



CIRCLES: 0 MILE RADIUS
MONITOR LOCATIONS:

▲ (6) NAMS
▽ (14) SLAMS
■ (8) OTHER
□ (0) INDUSTRIAL
★ (8) UNKNOWN
◆ (0) TRIBAL
✚ (0) INDEX SITE
* (0) NON-EPA

■ PLANT LOCATION
(516 DISPLAYED)
(6 NOT DISPLAYED,
LACK LAT.-LONG.)

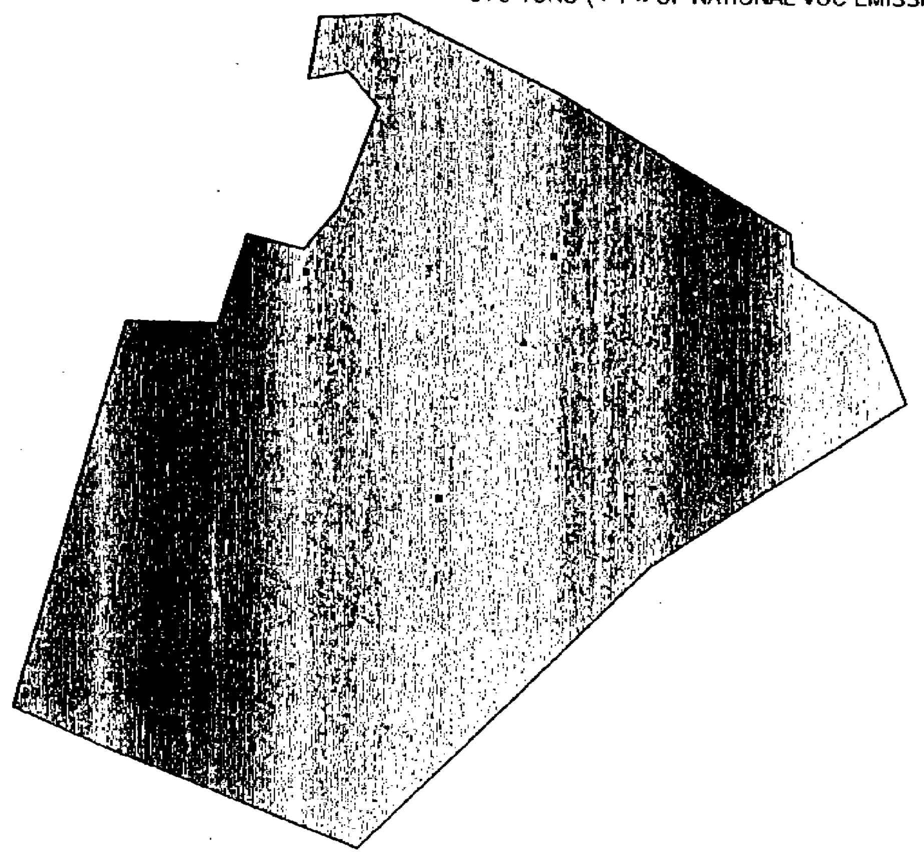
USEPA, OAQPS, TSD, NATIONAL AIR DATA BRANCH

08/13/93

WAKE CO, NORTH CAROLINA [illegible] MONITORS FOR O3 (ACTIVE [illegible]
(All) — ALL SIC CODES [illegible]
FACILITIES WITH VOC EMISSIONS [illegible] RY IN 1989 ((All))

SHADED COUNTIES HAVE MONITORS

670 TONS (< 1 % OF NATIONAL VOC EMISSIONS)



CIRCLES: 0 MILE RADIUS

MONITOR LOCATIONS: ▲ (2) NAMS ▽ (0) SLAMS ■ (1) OTHER □ (0) INDUSTRIAL
★ (0) UNKNOWN ◆ (0) TRIBAL + (0) INDEX SITE * (0) NON-EPA

■ PLANT LOCATION
(5 DISPLAYED)

USEPA, OAQPS, TSD, NATIONAL AIR DATA BRANCH

08/13/93

APPENDIX L

BIBLIOGRAPHY

BIBLIOGRAPHY

1. *AIRS Users Guide*, Volume AQ2, "Air Quality Data Coding," U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 1993.

2. *AIRS Users Guide*, Volume AQ3, "Air Quality Data Storage," U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 1993.

3. *AIRS Users Guide*, Volume AQ4, "Air Quality Data Retrieval," U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 1993.

4. *AIRS Users Guide*, Volume AQ5, "AIRS Ad Hoc Retrieval," U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 1993.

5. *Ambient Monitoring Guidelines for Prevention of Significant Deterioration (PSD)*, EPA-450/4-87-007, U.S. Environmental Protection Agency, Research Triangle Park, May 1987.

6. Berg, Neil J., et al., *Enhanced Ozone Monitoring Network Design and Siting Criteria Guidance Document*, EPA-450/4-91-033, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, November 1991.

7. *Clean Air Act Ozone Design Value Study*, Preliminary Draft, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, April 1993.

8. *Code of Federal Regulations*, Title 40, Part 50, U.S. Government Printing Office, 1992.

9. *Code of Federal Regulations*, Title 40, Part 51, U.S. Government Printing Office, 1992.

10. *Code of Federal Regulations*, Title 40, Part 53, U.S. Government Printing Office, 1992.

11. *Code of Federal Regulations*, Title 40, Part 58, U.S. Government Printing Office, 1992.

12. Cox, William M. and Shao-Hang Chu, "Meteorologically Adjusted Ozone Trends in Urban Areas: A Probabilistic Approach," *Tropospheric Ozone and the Environment II*, Air and Waste Management Association, Pittsburgh, Pennsylvania, 1992.

13. *Criteria For Assessing The Role of Transported Ozone/Precursors in Ozone Nonattainment Areas*, EPA-450/4-91-015, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, May 1991.

14. Dorosz-Stargardt, Geri, "Initial Implementation of the Photochemical Assessment Monitoring Stations (PAMS) Network," U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 1993.

15. *Federal Register*, (57 FR 7687), "Ambient Air Quality Surveillance - Proposed Rule," March 4, 1992.

16. *Federal Register*, (58 FR 8452), "Ambient Air Quality Surveillance - Final Rule," February 12, 1993.

17. Gerald, Nash O., William F. Hunt, Jr., Geri Dorosz-Stargardt, and Neil H. Frank, "Requirements for the Establishment of Enhanced Ozone Monitoring Networks," presented at the Air and Waste Management/EPA Symposium "Measurement of Toxic and Related Air Pollutants," Durham, North Carolina, May 4-7, 1993.

18. "Guidance for the Development and Approval of Photochemical Assessment Monitoring Stations Network Plans," Preliminary Draft, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, June 1993.

19. *Guideline for the Implementation of the Ambient Air Monitoring Regulations 40 CFR Part 58*, EPA-450/4-79-038, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, November 1979.

20. *Guideline for the Interpretation of Ozone Air Quality Standards*, EPA-450/4-79-003, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, January 1979.

21. "Guideline on Modification to Monitoring Seasons for Ozone," U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, March 1990.

22. Hunt, William F., Jr. and Nash O. Gerald, "The Enhanced Ozone Monitoring Network Required by the New Clean Air Act Amendments," 91-160.3, Air and Waste Management Association, Vancouver, 1991.

23. Koutrakis, Petros, Jack M. Wolfson, Arnold Bunyaviroch, Susan E. Froehlich, Koichiro Hirano, and James D. Mulik, "Measurement of Ambient Ozone Using a Nitrite-Coated Filter," *Analytical Chemistry*, Vol. 65, 1993.

24. Lewis, Charles W. and Teri L. Conner, "Source Reconciliation of Ambient Volatile Organic Compounds Measured in the Atlanta 1990 Summer Study: The Mobile Source Component," U.S. Environmental Protection Agency, Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, North Carolina, September 1991.

25. "List of Designated Reference and Equivalent Methods," U.S. Environmental Protection Agency, Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, North Carolina, November 12, 1993.

26. Liu, L.-J. Sally, Petros Koutrakis, Helen H. Suh, James D. Mulik, and Robert M. Burton, "Use of Personal Measurements for Ozone Exposure Assessment: A Pilot Study," "Environmental Health Perspectives," *Journal of the National Institute of Environmental Health Sciences*, Vol. 101, No. 4, September 1993.

27. Ludwig, F.L. and E. Shelar, *Site Selection for the Monitoring of Photochemical Air Pollutants*, EPA-450/3-78-013, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, April 1978.

28. McClenney, William A., "Instrumentation to Meet Requirements for Measurement of Ozone Precursor Hydrocarbons in the U.S.A.," U.S. Environmental Protection Agency, Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, North Carolina, 1993.

29. Mulik, James D., Jerry L. Varns, Petros Koutrakis, Mike Wolfson, Dennis Williams, William Ellenson, and Keith Kronmiller, "The Passive Sampling Device as a Simple Tool for Assessing Ecological Change - An Extended Monitoring Study in Ambient Air," presented at the Measurement of Toxic and Related Air Pollutants symposium, Durham, North Carolina, May 1992.

30. *On-Site Meteorological Instrumentation Requirements to Characterize Diffusion from Point Sources*, EPA-600/9-81-020, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 1981.

31. *On-Site Meteorological Program Guidance for Regulatory Modeling Applications*, EPA-450/4-87-013, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 1987.

32. "Ozone and Carbon Monoxide Areas Designated Nonattainment," U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, October 26, 1991.

33. Purdue, Larry J., "Continuous Monitoring of VOC Precursors," presented at the VOC Workshop Assessment and Evaluation, Amersfoort, The Netherlands, January 26-27, 1993.

34. Purdue, Larry J., "Summer 1990 Atlanta Ozone Precursor Study," presented at the VOC Workshop Assessment and Evaluation, Amersfoort, The Netherlands, January 26-28, 1993.

35. Purdue, Larry J., Dave-Paul Dayton, Joann Rice and Joan Bursey, *Technical Assistance Document for Sampling and Analysis of Ozone Precursors*, EPA-600/8-91-215, U.S. Environmental Protection Agency, Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, North Carolina, October 1991.

36. Purdue, Larry J., J. A. Reagan, W. A. Lonneman, T. C. Lawless, R. J. Drago, G. Zalaquet, M. Holdren, D. Smith, C. Spicer and A. Pate, *Atlanta Ozone Precursor Monitoring Study*, U.S. Environmental Protection Agency, Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, North Carolina, April 3, 1992.

37. *Quality Assurance Handbook for Air Pollution Measurement Systems*, Volume IV: Meteorological Measurements, EPA-600/4-82-060, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 1989.

38. *Rethinking The Ozone Problem In Urban And Regional Air Pollution*, National Research Council, National Academy Press, Washington, D.C., 1991.

39. Rice, Joann and Dave-Paul Dayton, *Enhanced Ozone Monitoring Methodology Evaluation*, Radian Corporation, Research Triangle Park, North Carolina, October 1992.

40. *The Role of Ozone Precursors in Tropospheric Ozone Formation and Control*, "A Report to Congress," EPA-454/R-93-024, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, July 1993.

41. Schweiss, Jon, "Draft Guidance for the Use of Portable Samplers," U.S. Environmental Protection Agency, Seattle, Washington, February 1991.

42. *Screening Procedures for Estimating the Air Quality Impact of Stationary Sources*, Revised, EPA-454/R-92-019, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, October 1992.

43. Shao-Hang Chu, "Meteorological Considerations in Siting Monitors of Photochemical Pollutants," presented at the Regional Photochemical Measurement and Modeling Study Conference, San Diego, California, November 1993.

44. *Technical Support for Enhanced Air Quality Modeling Analysis for the Purpose of Development of the 1994 Ozone State Implementation Plan Guidance*, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, Draft, April 1993.

TECHNICAL REPORT DATA *(Please read Instructions on the reverse before completing)*		
1. REPORT NO EPA-454/B-93-051	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE Photochemical Assessment Monitoring Stations Implementation Manual		5. REPORT DATE March 1994 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS TRC Environmental Corporation Chapel Hill, NC		10. PROGRAM ELEMENT NO. 11. CONTRACT/GRANT NO. EPA Contract 68D30029
12. SPONSORING AGENCY NAME AND ADDRESS U.S. Environmental Protection Agency Office of Air Quality Planning and Standards Technical Support Division, Monitoring & Reports Branch Research Triangle Par, NC 27711		13. TYPE OF REPORT AND PERIOD COVERED Final Report 14. SPONSORING AGENCY CODE
15. SUPPLEMENTARY NOTES EPA Work Assignment Manager: N. O. Gerald		

16. ABSTRACT

This document is designed to familiarize State and local air authorities with the Photochemical Assessment Monitoring Stations (PAMS) program and to provide guidance for designing PAMS monitoring networks. The document provides an explanation of the requirements of 40 CFR 58 pertaining to PAMS, specific guidance on network design and monitor siting, operational requirements, planning and approval processes, and data storage and communications systems.

17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Ambient Air Monitoring Photochemical Assessment Monitoring Stations (PAMS) Air Toxics Monitoring		
18. DISTRIBUTION STATEMENT Release Unlimited	19. SECURITY CLASS *(This Report)* Unclassified 20. SECURITY CLASS *(This page)* Unclassified	21. NO. OF PAGES 548 22. PRICE

EPA Form 2220-1 (9-73)

First
Class
Postage

U.S. Environmental Protection Agency
PAMS Implementation Manual Registration
Monitoring and Reports Branch (MD-14)
Research Triangle Park, NC 27711

Attn: Distribution Record System

PHOTOCHEMICAL ASSESSMENT MONITORING STATIONS (PAMS) IMPLEMENTATION MANUAL

DISTRIBUTION RECORD

A PAMS Manual distribution record system has been established and will be kept up to date so that future revisions to existing material and the addition of new material may be distributed to Manual users. In order to enter the ***Manual user's name and address in the distribution record system, the Pre-addressed "Distribution Record Card" must be filled out and returned.*** Any future changes in name/address should be sent to the following:

U.S. Environmental Protection Agency
PAMS Implementation Manual Registration
Monitoring and Reports Branch (MD-14)
Research Triangle Park, NC 27711

Attn: Distribution Record System

(cut along dotted line)

DISTRIBUTION RECORD CARD
(please print)

Date: ____

Last Name ____ First ____ Middle Initial ____

Agency, Company or Affiliate Name ____

Street ____

City ____ State ____ Zip Code ____

☐ I have received a copy of the PAMS Implementation Manual. Please send me any revisions and additions.

☐ Please send me the following specific revision and/or addition to the Manual. ____

☐ Please change my name and/or address on your distribution record system as reflected above.

Postmaster: Addressee other side

First
Class
Postage

U.S. Environmental Protection Agency
PAMS Implementation Manual Registration
Monitoring and Reports Branch (MD-14)
Research Triangle Park, NC 27711

Attn: Distribution Record System

PHOTOCHEMICAL ASSESSMENT MONITORING STATIONS (PAMS) IMPLEMENTATION MANUAL

DISTRIBUTION RECORD

A PAMS Manual distribution record system has been established and will be kept up to date so that future revisions to existing material and the addition of new material may be distributed to Manual users. In order to enter the ***Manual user's name and address in the distribution record system, the Pre-addressed "Distribution Record Card" must be filled out and returned.*** Any future changes in name/address should be sent to the following:

U.S. Environmental Protection Agency
PAMS Implementation Manual Registration
Monitoring and Reports Branch (MD-14)
Research Triangle Park, NC 27711

Attn: Distribution Record System

(cut along dotted line)

DISTRIBUTION RECORD CARD
(please print)

Date: ___

Last Name ___ First ___ Middle Initial

Agency, Company or Affiliate Name

Street

City ___ State ___ Zip Code

☐ I have received a copy of the PAMS Implementation Manual. Please send me any revisions and additions.

☐ Please send me the following specific revision and/or addition to the Manual. ___

☐ Please change my name and/or address on your distribution record system as reflected above.

Postmaster: Addressee other side

First
Class
Postage

U.S. Environmental Protection Agency
PAMS Implementation Manual Registration
Monitoring and Reports Branch (MD-14)
Research Triangle Park, NC 27711

Attn: Distribution Record System

PHOTOCHEMICAL ASSESSMENT MONITORING STATIONS (PAMS) IMPLEMENTATION MANUAL

DISTRIBUTION RECORD

A PAMS Manual distribution record system has been established and will be kept up to date so that future revisions to existing material and the addition of new material may be distributed to Manual users. In order to enter the ***Manual user's name and address in the distribution record system, the Pre-addressed "Distribution Record Card" must be filled out and returned.*** Any future changes in name/address should be sent to the following:

U.S. Environmental Protection Agency
PAMS Implementation Manual Registration
Monitoring and Reports Branch (MD-14)
Research Triangle Park, NC 27711

Attn: Distribution Record System

(cut along dotted line)

DISTRIBUTION RECORD CARD
(please print)

Date: ____

Last Name ____ First ____ Middle Initial ____

Agency, Company or Affiliate Name ____

Street ____

City ____ State ____ Zip Code ____

❒ I have received a copy of the PAMS Implementation Manual. Please send me any revisions and additions.

❒ Please send me the following specific revision and/or addition to the Manual. ____

❒ Please change my name and/or address on your distribution record system as reflected above.

Postmaster: Addressee other side